

STIC-Biotech/ChemLib

58562

From: Davis, Natalie
Sent: Thursday, January 17, 2002 2:55 PM
To: STIC-Biotech/ChemLib
Subject: 09/823101

Please search for the protein encoded by SEQ ID NO: 1-4

Thanks
Natalie A. Davis, PhD
Patent Examiner
Art Unit 1642
CM1, Rm 8B13
Mailbox 8E12
Ph (703) 308-6410

RECEIVED
JAN 17 2002
STIC

Searcher: B. Keane - Early
Phone: 308-4501
Location: Biotech Lab.
Date Picked Up: 1/17/02
Date Completed: 1/18/02
Searcher Prep/Review: _____
Clerical: 6 min
Online time: 3 min

TYPE OF SEARCH:
NA Sequences: _____
AA Sequences: 4
Structures: _____
Bibliographic: _____
Litigation: _____
Full text: _____
Patent Family: _____
Other: _____

VENDOR/COST(where applic.)
STN: _____
DIALOG: _____
Questel/Orbit: _____
DRLink: _____
Lexis/Nexis: _____
Sequence Sys.: AB5502
WWW/Internet: _____
Other (specify): _____

THIS PAGE BLANK (USPTO)


```

|||||  |||  ::|  |||||
75 ProProProProLeuProProSerPhePheLeuProProProLeuSe 91
546 ..... 546
91 rProLeuLeuLeuLeuPheLeuProPhePheLeuLeuLeuPheP 108
545 .....TTCCTTAATTTTGTTCCTACTCTCTCT 516
108 heProLeuLeuProLeuPhePheLeuPhePheLeuLeuLeuPro 124
515 TATTCCTCC.....CCCTTAAATAA 493
125 SerCysSerSerSerSerLeuLeuLeuLeuPheLeuProPhe...SerSe 140
492 ATTACTCTCTCTTTTCATATTGTATCTCCCTCGCTAATAAAAAA 443
140 rLeuSerSerSerProSerPheSerPheSerPheSerPheSerSers 157
442 AAATAACACCGCATGTGTTTGTATTTCTGTTTCGTTATCCATC 393
157 erPheProPheLeuLeuProPhePheLeuLeuSerLeuPheLeu 173
392 TTCCTGTGATATTTCCTGTCCTCCGCTCTCCCTCCCGCATAT 343
174 PheLeuLeuSerLeuLeuCysPheSerLeuLeuProProProPheph 190
342 AAATATCTCTGTTATTGA 324
190 eLeuLeuLeuLeuLeu 196

seq_name: /SID8/gcgdata/geneseq/geneseq/AA2001.DAT:AA28950
seq_documentation_block:
ID AA28950 standard; Protein; 242 AA.
XX
XX AA28950;
XX
XX 17-OCT-2001 (first entry)
XX
DE Peptide #2987 encoded by probe for measuring placental gene expression.
XX
XX Probe: microarray; human; placenta; antenatal diagnosis;
XX
XX genetic disorder.
XX
OS Homo sapiens.
XX
XX WO200157272-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00663.
XX
XX 04-FEB-2000; 2000US-0180312.
XX 26-MAY-2000; 2000US-0207456.
XX 30-JUN-2000; 2000US-0608408.
XX 03-AUG-2000; 2000US-0632366.
XX 21-SEP-2000; 2000US-0234687.
XX 27-SEP-2000; 2000US-0236359.
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI, 2001-488897/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human placenta -
XX
XX Claim 27; SEQ ID No 29219; 654bp; English.
XX
XX The present invention relates to single exon nucleic acid probes (SENPs:

```

```

CC see AA131315-AA157546). The present sequence is a peptide encoded by one
CC such probe. The probes are useful for producing a microarray for
CC predicting, measuring and displaying gene expression in samples derived
CC from human placenta. The probes are useful for antenatal diagnosis of
CC human genetic disorders.
XX
XX Sequence 242 AA:

Alignment_scores:
Quality: 92.50 Length: 123
Ratio: 1.745 Gaps:
Percent Similarity: 43.089 Percent Identity: 26.829

Alignment_block:
US-09-823-101-3/rev x AA28950 ..

Align seg 1/1 to: AA28950 from: 1 to: 242

578 CCTCTCCCTCCCTCCGCAAAATAAGTTTGTG..... 546
|||||  |||  ::|  |||||
75 ProProProProLeuProProSerPhePheLeuProProProLeuSe 91
546 ..... 546
91 rProLeuLeuLeuLeuPheLeuProPhePheLeuLeuLeuPheP 108
545 .....TTCCTTAATTTTGTTCCTACTCTCTCT 516
108 heProLeuLeuProLeuPhePheLeuPhePheLeuLeuLeuPro 124
515 TATTCCTCC.....CCCTTAAATAA 493
125 SerCysSerSerSerSerLeuLeuLeuLeuPheLeuProPhe...SerSe 140
492 ATTACTCTCTCTTTTCATATTGTATCTCCCTCGCTAATAAAAAA 443
140 rLeuSerSerSerProSerPheSerPheSerPheSerPheSerSers 157
442 AAATAACACCGCATGTGTTTGTATTTCTGTTTCGTTATCCATC 393
157 erPheProPheLeuLeuProPhePheLeuLeuSerLeuPheLeu 173
392 TTCCTGTGATATTTCCTGTCCTCCGCTCTCCCTCCCGCATAT 343
174 PheLeuLeuSerLeuLeuCysPheSerLeuLeuProProProPheph 190
342 AAATATCTCTGTTATTGA 324
190 eLeuLeuLeuLeuLeu 196

seq_name: /SID8/gcgdata/geneseq/geneseq/AA2001.DAT:AA04184
seq_documentation_block:
ID AA04184 standard; Protein; 242 AA.
XX
XX AA04184;
XX
XX 09-OCT-2001 (first entry)
XX
XX Peptide #2866 encoded by probe for measuring breast gene expression.
XX
XX Probe: human; breast disease; breast cancer; development disorder;
XX inflammatory disease; proliferative breast disease; non-carcinoma tumour.
XX
XX Homo sapiens.
XX
XX WO200157270-A2.
XX
XX 09-AUG-2001.
XX
XX 29-JAN-2001; 2001WO-US00661.
XX

```



```

72 scys..... 73
161 TTCACGCACTTGCTGTCATGCATATGACTGTGAGATTATAG 112
74 .....ArgMetCysGlnTyrValIleGluCysArgValGlu.AlaAlaG1 88
111 ATACGCTAGACACGATGATGACAGACTCACAGAGATCCAGCA 62
88 YTYrPheArGThrPheTyrGlyLysArgPheAsnPhcGlnGluProGlyL 105
61 AAGCAGACCTGTGACAGCATAGGTCACACAGACAGACCCAGTGACA 12
105 ySTYrValIleuAlaArgGly.....ThrySGlyGlyAspTrpSer 118

seq_name: /SID8/gcgdata/geneseq/AA198.DAT.AAM59825
seq_documentation_block:
ID AAM59825 standard; Protein; 555 AA.
XX
AC AAM59825;
XX
XX 07-DEC-1998 (first entry)
XX
DE Cypridina luciferase.
XX
KW Luciferase; reporter; analysis; exocytosis;
KW cellular microenvironment; optical label.
XX
OS Cypridina hllegendorffii.
XX
PN W09836081-A2.
XX
PD 20-AUG-1998.
XX
PF 13-FEB-1998; 98WO-US02774.
XX
PR 14-FEB-1997; 97US-0036805.
PR 13-FEB-1997; 97US-0038179.
XX
XX (SLOK ) SLOAN KETTERING INST CANCER RES.
XX
PI De Angelis DA, Miesenbock G, Rothman JE;
XX
DR WPI; 1998-457118/39.
DR N-PSDB; AAV41716.
XX
XX New polypeptide for monitoring changes in molecular environment,
XX especially release of synaptic vesicles - comprises targeting
XX component and reporter producing optical signal in contact with
XX second compartment, also new pH-sensitive mutants of green
XX fluorescent protein and related nucleic acid
XX
PS Claim 10; Fig 13; 127pp; English.
XX
XX This is the amino acid sequence of Cypridina hllegendorffii
XX luciferase. The invention relates to methods and compositions
XX which utilise the emission of light to monitor changes in cellular
XX microenvironments. It is especially useful for monitoring
XX exocytotic activity such as quantal release of synaptic vesicles.
XX Fusion proteins of Cypridina luciferase and synaptobrevin-1 or
XX VAMP/synapobrevin-2 were targeted to synaptic vesicles and, upon
XX exocytosis, formed light-emitting complexes with luciferin present
XX in the extracellular medium. Photon emissions in the presence of
XX a depolarising stimulus can be observed with these systems. The
XX invention allows analysis of cells, cultures, tissue sections etc.
XX
SQ Sequence 555 AA;

```

```

alignment_scores:
Quality: 90.50 Length: 134
Ratio: 1.775 Gaps: 6

```

```

Percent Similarity: 38.060 Percent Identity: 28.358
alignment_block:
US-09-823-101-3/rev x AAM59825
Align seq 1/1 to: AAM59825 from: 1 to: 555
372 TTGCCGCTCTCTCCGCCCCCTCATATATAATATCTTTATTATA 323
22 CysProValGluAlaGluProPheSer..... 30
322 TTGGTGGGGTGATTAATTTTGGGGGAAAACTCATGTCGCCCATCGG 273
31 .....SerThrProThrValProThrSerC 39
272 GGATAGCTCCAC.....ANGTCAGCTGTGCTATTA..... 241
39 ysgIuAlaLysGluGlyGluCysIleAspThrArgCysAlaThrCysLys 55
240 .....CTCTGTCAAAACCAACANTGTGACACACAG 212
56 ArgAspIleuSerAspGlyLeuCysGluAsnLysProGlyLysThrC 72
211 TTGTGCTCTCATGTTGTTGACACGTCATGTCGACTGCATGCCACAG 162
72 scys..... 73
161 TTCACGCACTTGCTGTCATGCATATGACTGTGAGATTATAG 112
74 .....ArgMetCysGlnTyrValIleGluCysArgValGlu.AlaAlaG1 88
111 ATACGCTAGACACGATGATGACAGACTCACAGAGATCCAGCA 62
88 YTYrPheArGThrPheTyrGlyLysArgPheAsnPhcGlnGluProGlyL 105
61 AAGCAGACCTGTGTCACAGCATAGGTCACACAGACAGACCCAGTGACA 12
105 ySTYrValIleuAlaArgGly.....ThrySGlyGlyAspTrpSer 118

seq_name: /SID8/gcgdata/geneseq/AA2001.DAT.AAB51162
seq_documentation_block:
ID AAB51162 standard; Protein; 555 AA.
XX
AC AAB51162;
XX
XX 21-MAR-2001 (first entry)
XX
DE Vargula (Cypridina) luciferase protein sequence SEQ ID NO:4.
XX
XX Bioluminescence generating system; fluorescent protein; recreation;
XX entertainment; amusement; toy; paint; slimy play material; textile;
XX bubble making toy; balloon; cosmetic; bath powder; body lotion; gel;
XX powder; cream; toothpaste; dentifrice; soap; body paint; bubble bath;
XX gelatin; icing; frosting; beverage; beer; wine; champagne; soft drink;
XX glowing ice; fountain; liquid firework.
XX
OS Vargula hllegendorffii.
XX
PN US6152358-A.
XX
PD 28-NOV-2000.
XX
PF 17-AUG-1998; 98US-0135988.
XX
PR 06-FEB-1996; 96US-0597274.
PR 25-NOV-1996; 96US-0757046.
XX
XX (BRYA/) BRYAN B.
XX
PI Bryan B;
XX
DR WPI; 2001-060056/07.

```

DR N-PSDB; AAC92232.

XX New combinations of articles of manufacture and with bioluminescence
PT generating systems, useful for producing novelty items for
PT entertainment, recreation and amusement, personal items or foods, e.g.
PT toys, cosmetics or gelatins

PS Disclosure: Column 105-110; 83pp; English.

XX The present invention describes a combination comprising an article of
CC manufacture and one or more components of a bioluminescence generating
CC system. The combination is a novelty item and the article of manufacture
CC is a paper product. The combination is useful for producing novelty
CC items, particularly those that glow, or produce or expel a
CC bioluminescent composition. The novelty items are useful for
CC entertainment, recreation and amusement, personal items, or foods.
CC Specifically, these include toys, paints, slimy play materials, textiles
CC (e.g. clothing), bubbles in bubble making toys and other toys that
CC produce bubbles, balloons, cosmetics, bath powders, body lotions, gels,
CC powders and creams, toothpastes and other dentifrices, soaps, body
CC paints, or bubble baths, gelatins, ictings and frostings, beverages (beer,
CC wine, champagne, or soft drinks), glowing ice, fountains (e.g. liquid
CC fireworks) and other jets or sprays or aerosols of compositions that are
CC solutions, mixtures, suspensions, powders, pastes, or particles. AAC92229
CC to AAC92241 and AAB51159 to AAB51172 represent sequences which are given
CC in the exemplification of the present invention.

XX Sequence 555 AA;

alignment_scores:
Quality: 90.50 Length: 134
Ratio: 1.775 Gaps: 6
Percent Similarity: 38.060 Percent Identity: 28.358

alignment_block:

US-09-823-101-3/rev x AAB51162 ..

Align seg 1/1 to: AAB51162 from: 1 to: 555

```
372 TGTCCGCTCTCTCCGCCCTCATATAAATAATCTCTTTATTATTA 323
      |||||
22 CysProvalaGluAlaGluProSer..... 30
322 TTGGTGGGGGTGATTAATTTTGGGGGAAACATCATGCCCCCATCGG 273
      |||||
31 .....SerThrProThrValProThrSerC 39
```

```
272 GGATAGCTCCNCAC.....ANGTCAGTTGTGCTATA..... 241
      |||
39 ysgluAlaIeLysgluIuGlycysIleAspThrArgCysAlaThrCysLys 55
240 .....CTCTGTTCAAAACANTGTGGACACACAG 212
      |||||
```

56 ArgAspIleLeuSerAspIleuGlyLeuGlyCysIleAsnLysProGlyLysThrCys 72

211 TTGTGCTCATGTGGTGGACACGTCATCGTGAACATGCATGCCACAG 162

72 scYs..... 73

161 TTCACGACGACGCTTGTCTGTCATGATGATGACTGTGAGATTATAGG 112

74ArgMetCysGlnIrrValIleGluCysArgValGluAlaIle 88

111 ATACGCTAGACAGGATGATAGACAGATCACAAGAGATCCAGGCA 62

88 yIrrPheArgThrPheTyrGlyLysArgPheAsnIleGluProGlyL 105

61 AAGCACACCTGGTGAACAGCATAGGTCACACAGAGACAGCCAGGAGACA 12

105 ystYrrValIleuAlaArgIly.....ThrLysGlyIlyAspIrrSer 118

seq_name: /SID58/ycgdata/geneseq/AA1988.DAT: AAP80656

seq_documentation_block:

ID AAP80656 standard; protein; 562 AA.

XX AAP80656;

XX 07-OCR-1990 (first entry)

DE Tissue plasminogen activator analogue.

KW Tissue plasminogen activator; analogue; K1 domain.

PN EP293934-A.

PD 07-DEC-1988.

PF 03-JUN-1988; 88EP-0108949.

PR 04-JUN-1987; 87US-0058217.

PA (EISA-) EISAI CO. LTD.

PA (ZYMO-) ZYMOGENETICS INC.

PA (NOVO-) NOVO INDUSTRI A/S.

PI Mulvihill ER, Nexo BA, Yoshitake S, Ikeda A, Suzuki S, Hashimoto A;

PI yazuritha T;

DR WPI: 1988-347625/49.

DR N-PSDB; AAN81091.

XX New tissue plasminogen activator analogues with non-native K1-domain -

XX having higher fibrin specificity, and new encoding DNA sequences.

XX Disclosure: 4 App; English.

XX The polypeptide is encoded by a mutated t-PA sequence in plasmid
CC Zm99-9200. The t-PA analogue has the K1 domain of native t-PA replaced
CC by another kringle domain, esp. the K1 domain of plasminogen, which
CC mediates the binding of the analogue to fibrin and contains 6 Cys.
CC The analogue has greater specificity for fibrin than native t-PA. It is a
CC useful fibrinolytic agent with specificity for clot lysis. The analogue
CC can be tailored to increase plasma half life, solubility etc. It may also
CC include a finger, or growth factor domain of native t-PA, protein C
CC or factor VII.

XX Sequence 562 AA;

alignment_scores:
Quality: 83.50 Length: 171
Ratio: 1.210 Gaps: 11
Percent Similarity: 40.351 Percent Identity: 24.561

alignment_block:

US-09-823-101-3 x AAP80656 ..

Align seg 1/1 to: AAP80656 from: 1 to: 562

```
40 ATGCCCTGACACGAGTGTGCTTGCCT.....GGATCTTCTTG 77
      |||
81 ValProvalaLysSerCysSerIuProThrCysPheAsnGlyGlyThrCys 97
78 TGAGTCTCGCTA..... 90

```

97 scIuGlnAlaIeUrrPheSerAspPheValIysGlnCysProGluGlyP 114

91TGCAATCCCTGCTAGCGTATCTATTAATCTCAACAGT 129

114 heAlaGlyLysCys.....Ser 119

130 CATATGATGACACAGACAGCAAGTGCCTGGAACGCT.....G 167

120 GluIleAspThr.....ArgIleThrCysTyrGluAspGln 132

```

168 GCATGACGTCACGATGACGTGTCACCAACATAGACACAACCTGTG 217
132 LyleSerTyArgGlyThrTrpSerThrAlaGluArgAla...AlaGlu 147
218 TGTCACACNTGTTGAACAGAGATAGACACAGCTGACNTGTGNGAGC 267
148 CysThrAsnTrpAsnSerSerAlaLeuAlaGlnLysProTyrSerGlyAr 164
268 TATCCCGATGGGGGACATGATTTTCCCAAAAATTTATCCACCCC 317
164 GARGProAspAla..... 168
318 ACCAATTAAATACAGAGATTTTATATGAGGGGGGAGAGAGACG 367
169 .....IleArgLeuGlyLeuGlyAsnHisAsn 177
368 GGACACAGGAATATCAACAGAGATGATTAACGAAAACAGAAATAC 417
178 AspCysArg.....AsnProAspArgAspSe 186
418 AAAAACAACATGCGGTATTTTATTTTATAGCAGAGAGAT 467
186 rLys.....AlaTrpCysTyrValPhe.....LysGlyGlyLysT 198
468 ACAATATATGAA 478
198 YrSerSerGlu 201

```

seq_name: /SIDS8/gcgdata/geneseq/geneseqp/AA2001.DAT:AA672099

seq_documentation_block:

ID AAG72099 standard; Protein: 328 AA.

AC AAG72099;

DT 31-JUL-2001 (first entry)

DE Human olfactory receptor polypeptide, SEQ ID NO: 1780.

XX Human: olfactory receptor; OR: primary scent determination;

KW secondary scent determination; polypeptide library; odour receptor;

KW scent profile; scent fingerprint; scent representation.

OS Homo sapiens.

PN WO200127158-A2.

PD 19-APR-2001.

PF 06-OCT-2000; 2000WO-US27582.

XX 08-OCT-1999; 99US-0158615.

PR 24-FEB-2000; 2000US-0184809.

PA (DIGI-) DIGISCENTS.

PA (YEDA) YEDA RES & DEV CO LTD.

PI Bellenson J, Smith D, Lancet D, Glusman G, Fuchs T, Yanai I;

DR WPI: 2001-290713/30.

XX New polynucleotides which encode polypeptides involved in olfactory

PT sensation for identifying olfactory agonists and antagonists -

XX Claim 11; Page 1169-1170; 1857pp; English.

CC The present sequence is an olfactory receptor which is encoded by

CC one of a number of novel polynucleotides. The polynucleotides can be

CC used in screening for olfactory agonists and antagonists. The methods

CC allow for the determination of primary scents and the identification

CC of the odour receptors used to detect these primary scents. The methods

CC also enable determination of secondary scents and the identification of

CC combinations of odour receptors that are involved in detecting such
 CC secondary scents. This enables the construction of a scent representation
 CC (also called a scent fingerprint or scent profile), which may be used to
 CC re-create and edit scents. Libraries of olfactory receptors are useful
 CC for determining the interaction pattern of a composition with the
 CC receptors, and can be used for determining differences in the olfactory
 CC faculties of different individuals.

SO Sequence 328 AA;

alignment_scores: Quality: 83.00 Length: 196

Percent Similarity: 45.408 Gaps: 8

Percent Identity: 24.490

alignment_block:

US-09-823-101-3/rev x AAG72099 ..

Align seq 1/1 to: AAG72099 from: 1 to: 328

```

680 TATTTTGTTCATATAGTCGCCCCCACTAATATTTCTCGG 631
28 PhePheLeuPheLeuValIleTyrValValThrLeuLeuGlyAsnSerHisL 44
630 CGTCCCTCTCATAGCTGCATATATATATATTTCTCTGCTG 581
44 yLeuVal.....ThrIleGlyLeuAsnSerHisL 55
580 TCCCTCTCTCCCCCACCACAAAAATAGTTTGTGTCTTAAATTTTGT 531
55 eunIsthrPro.....MetyrPhePheLeuPheAsnLeuSer 67
530 TTCTCTCTCTCTCTTATTCCTCTCTCTCTCTCTCTCTCT 481
68 PheIleAspLeuCysTyrSerSerValPheThrProLysMetLeuMetAs,84
480 TTTTCATATTTTGAATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 431
84 nPhe.....IleSerGlyLysAsnIleIleSerPheL 95
430 GCATGTCTTT.....TTTGTATTTCTGTCTTTGTTATTCATCTT 391
95 ysgLysMetThrGlnLeuPhePheLeuPheLeuValIleSer 111
390 CTTGTGTGATATTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 355
112 GlucYs...TyrValLeuThrSerMetAlaTyrAspArgValAlaIleCy 127
354 .....CCCCCTCATATATAATATCTCTTTTATTAATTTGGGGTGGAT 309
127 sasnProLeuLeuTyrHisIleAlaMetSerProThrValCysSerSerL 144
308 AAATTTTGGGGGAAAAACTCATGTCCCCCATCGGGGATAGCTCCNCAC 259
144 eumetPheGlySer..... 148
258 ANGTCACGTTGTGCTATACCTGTTCACCAACCAANTGGACACACAAGTG 209
149 .....TyrLeuMetAlaPheSerGlyAlaMetAlaHisThrGlyCy 162
208 TGCTCTCATGTG.....GTTGACCACTGATCG 180
162 smetLeuArgLeuThrPheCysAspAlaAsnThrIleAspHisIstTyrPhec 179
179 TGAACCTGCATGCCACACGTTCCACGACCTTGCTGTCT 142
179 yAspIleLeuProLeuLeuGlnLeuSerCysThrSer 191

```

seq_name: /SIDS8/gcgdata/geneseq/geneseqp/AA1990.DAT:AA05434

seq_documentation_block:

ID AAR05434 standard; protein: 555 AA.

```
XX AC AAR05434;
XX DF 30-JUL-1990 (first entry)
XX DE Sequence for luciferase.
XX KW Luciferase; ss.
XX OS Cypridina hilgendorffii ('umihotaru').
XX PN W09001542-A.
XX PD 22-FEB-1990.
XX PF 09-AUG-1989; 89WO-JP00811.
XX PR 17-AUG-1988; 88JP-0204173.
XX PA (TORA ) TORAY IND INC.
XX PI Kazaml J, Nakamura H, Goto T;
XX DR WPI; 1990-083501/11.
XX DR N-PSDB; AAQ02303.
XX PT Recombinant Luciferase, fragments from it, and gene coding for it -
XX PT the luciferase having increased stability and quantum yield.
XX PS Claim: Fig 1A-1D; 56pp; Japanese.
XX CC A recombinant vector wherein the luciferase-coding gene is connected
XX CC to the downstream portion of a promoter capable of expressing in a host
XX CC cell allows for the production of luciferase, having increased stability,
XX CC better quantum yield and light production conditions. Light of different
XX CC wavelengths can be obtained.
XX SQ Sequence 555 AA;

alignment_scores:
    Quality: 82.50      Length: 134
    Ratio: 1.650       Gaps: 6
    Percent Similarity: 37.313    Percent Identity: 27.612

alignment_block:
US-09-823-101-3/rev x AAR05434 ..

Align seg 1/1 to: AAR05434 from: 1 to: 555

372 TGTCCGCTCTCTCCGCCGCCCTCATATAAATATCTCTGTATTATA 323
    |||||
22 CyspPvovAlGluAlaGluAlaProSer..... 30
322 TTGGTGGGGGTGATTAATTTTGGGGGAAAACTCATGTCCCCCATCGG 273
    ::|||
31 .....SerThrProThrValProThrSerc 39
272 GGATGACTCCNCAC.....ANGTCACGTTGCTGTATA..... 241
    |||:::|||||
39 ysgluAlaLysgluLysgluLysleuAspThrArgCysAlaThrCysLys 55
240 .....CTCTGTCAAACCAANTGTGGACACACAG 212
56 ArgAspIleLeuSerAspLysLeuCysGluAsnLysProGlyLysThrCys 72
211 TTGTGCTCTCATGTGGTGGACACAGCTCATCGTGAACCTCATGCCACAG 162
    |||
72 scys..... 73

161 TTTCACAGCATGCTGTCTGCATGATATGACTGTAGAAATTATAG 112
    ||| ||| |||||
74 .....ArgMetCysGlnTyrValIleGluCysArgValGluAlaAlaG1 88
```

```
111 ATACGCTAGACAGAGATGCATAGACAGACTTCACAGAAGATCCAGCA 62
    ||| ||||| ::||| ::| ::|||
88 yTyrPheArgThrPheTyrGlyLysArgPheAsnPheGlnGluProGlyL 105
61 AAGCACACCTGCTGACGATAGGGTCCACAGACAGAAACCACTGGACA 12
    || ||||| ||| ||| ::|||
105 ystYrValLeuAlaArgGly.....ThrLysGlyLysAspTrpSer 118

seq_name: /SISB8/gcdata/geneseq/geneseq/AA2001.DAT: AAB51160
seq_documentation_block:
ID AAB51160 standard; Protein; 555 AA.
XX AC AAB51160;
XX DF 21-MAR-2001 (first entry)
XX DE Cypridina hilgendorffii luciferase protein sequence SEQ ID NO:2.
XX KW Bioluminescence generating system; fluorescent protein; recreation;
XX KW entertainment; amusement; toy; paint; slimy play material; textile;
XX KW bubble making toy; balloon; cosmetic; bath powder; body lotion; gel;
XX KW powder; cream; toothpaste; dentifrice; soap; body paint; bubble bath;
XX KW gelatin; icing; frosting; beverage; beer; wine; champagne; soft drink;
XX KW glowing ice; fountain; liquid firework.
XX OS Vargula hilgendorffii.
XX PN US6152358-A.
XX PD 28-NOV-2000.
XX PF 17-AUG-1998; 98US-0135988.
XX PR 06-FEB-1996; 96US-0597274.
XX PR 25-NOV-1996; 96US-0757046.
XX PA (BRYA/) BRYAN B.
XX PI Bryan B;
XX DR WPI; 2001-060056/07.
XX DR N-PSDB; AAC92230.
XX PT New combinations of articles of manufacture and with bioluminescence
XX PT generating systems, useful for producing novelty items for
XX PT entertainment, recreation and amusement, personal items or foods, e.g.
XX PT toys, cosmetics or gelatins -
XX PS Disclosure: Column 97-100; 83pp; English.
XX CC The present invention describes a combination comprising an article of
XX CC manufacture and one or more components of a bioluminescence generating
XX CC system. The combination is a novelty item and the article of manufacture
XX CC is a paper product. The combination is useful for producing novelty
XX CC items, particularly those that glow, or produce or expel a
XX CC bioluminescent composition. The novelty items are useful for
XX CC entertainment, recreation and amusement, personal items, or foods.
XX CC Specifically, these include toys, paints, slimy play materials, textiles
XX CC (e.g. clothing), bubbles in bubble making toys and other toys that
XX CC produce bubbles, balloons, cosmetics, bath powders, body lotions, gels,
XX CC powders and creams, toothpastes and other dentifrices, soaps, body
XX CC paints, or bubble baths, gelatins, icings and frostings, beverages (beer,
XX CC wine, champagne, or soft drinks), glowing ice, fountains (e.g. liquid
XX CC fireworks) and other jets or sprays or aerosols of compositions that are
XX CC solutions, mixtures, suspensions, powders, pastes, or particles. AAC92229
XX CC to AAC92241 and AAB51159 to AAB51172 represent sequences which are given
XX CC in the exemplification of the present invention.
XX SQ Sequence 555 AA;
```

alignment_scores: Length: 134
 Quality: 82.50
 Ratio: 1.650
 Gaps: 6
 Percent Similarity: 37.313 Percent Identity: 27.612

alignment_block:

US-09-823-101-3/rev x AAB51160 ..

Align seg 1/1 to: AAB51160 from: 1 to: 555

```

372 TGTCCCGCTCTCTCCCGCCCGCTCATATAAATATCTTGTATTAA 323
    |||||
22 CysProValGluAlaGluAlaProSer..... 30
322 TTGGTGGGCGTGAATAATTTTGGGGGAAAAACATCATGCCCGCATCGG 273
    |||||
31 .....SerThrProThrValProThrSerc 39
    |||||
272 GGATAGCTCCNCAC.....ANGTCACGTTGTGTATA..... 241
    |||||
39 ysgluAlaLysgluLysgluLysleasPThrArgCysAlaThrCysLys 55
    |||||
240 .....CTCTGTCAACCAACANTGTGCACACACAG 212
    |||||
56 ArgAspIleuSerAspGlyLeuLysCysGluAsnLysProGlyLysThrCys 72
211 TTGTGCTCTCATGTTGGTTGACACAGTCATCGTGAACGTCACACAG 162
    |||
72 scys..... 73
161 TTCACGCACTTGCTGTCTGTGCATGATATGATGTTGAGAAATTATAG 112
    |||||
74 .....ArgMetCysGlnTyValIleGlnSerArgValGlu..AlaAlaGlu 88
111 ATAGCGTAGCCAGCATGATGATGATGATGATGATGATGATGATGATGAT 62
    |||||
88 YTYrPheArgThrPheTyArgValAlaLysArgPheAsnPheGlnGluProGlyL 105
61 AAGCACACCTGTGTGACAGCATAGGTCACAGACAGAACCCAGTGGACA 12
    |||||
105 ystyValleuAlaLysArgGly.....ThrLysGlyGlyAspTrpSer 118

```

seq_name: /SID58/gcgdata/geneseq/geneseqp/AA1997.DAT.AAM36104

seq_documentation_block:

ID AAM36104 standard; Protein: 597 AA.

AAM36104:

25-MAR-1998 (first entry)

Luciferase/epidermal growth factor receptor fusion protein.

Sea firefly; Vargula sp.; Luciferase; label; mouse; murine; epidermal growth factor receptor; luminescent enzyme; fusion protein.

Chimeric - Vargula sp.

Chimeric - Mus sp.

Location/Qualifiers

Key

Domain

Region

Domain

Domain

Domain

Domain

Domain

Domain

Domain

Domain

Domain

Domain

Domain

Domain

Domain

Domain

DE D. melanogaster glutamate transporter (I) sequence.
 XX
 KW Glutamate transporter; nervous system; research reagent;
 KW arthropod; Drosophila melanogaster.
 XX
 OS Drosophila melanogaster.
 PN JP1009083-A.
 XX
 PD 21-APR-1998.
 XX
 PF 27-SEP-1996; 96JP-0290990.
 XX
 PR 27-SEP-1996; 96JP-0290990.
 XX
 PA (SUNR) SUNTOR LTD.
 XX
 DR WPI: 1998-289876/26.
 DR N-PSDB; AAV37697.
 XX
 PT Glutamate transporter obtained from arthropod - useful as reagent in
 XX research on nervous system
 PS
 PS Claim 7; Pages 5-7; 15pp; Japanese.
 XX
 CC This represents a Drosophila melanogaster glutamate transporter. The
 CC specification provides glutamate transporter genes from an arthropod,
 CC as shown in AAV37697 and AAV37698 and the corresponding protein
 CC sequences are shown in AAW61910 and AAW61911. A glutamate transporter
 CC gene encoding a glutamate transporter containing an amino acid sequence
 CC in which at least 1 amino acid is added, deleted or replaced in the amino
 CC acid sequence encoded by the above genes can be used in a method for
 CC screening a glutamate transporter inhibitor where the glutamate
 CC transporter gene is expressed in a culture cell and its glutamate
 CC transporter activity is measured. The transporter can be used as a
 CC reagent for the research on the nervous system.
 XX
 XX Sequence 473 AA:
 50
 alignment_scores:
 Quality: 81.50 Length: 130
 Ratio: 1.216 Gaps: 5
 Percent Similarity: 51.538 Percent Identity: 28.462
 alignment_block:
 US-09-823-101-3/rev x AAW61910 ..
 Align seg 1/1 to: AAW61910 from: 1 to: 473
 653 CCCCCCAATATATTTCTCGCGCCCTCATAGCTGCATATATAC 604
 |||||
 17 ProProteinSerAlcysLeuValaspSerSerSerSerSer 33
 603 TATATATATCTCTCG.....TCGTCTCCCTCCCTCC 569
 |||
 33 sileAlaLeuAlaSerGlyArgArgLuarSerGly.....Thrtyp 48
 568 CCCCACCAAAAAATAGTTTGTCTTAAATTTGTTTCTACCTCT 519
 |||
 48 roserProAlaLysIlePheLeuArgMetLeuLysCysLeuVal... 63
 518 CCTATATCCCTCCCTTAAATATATATCTCTCTTTCATATTTG 469
 |||
 64ProLeuLeuValSerSerIleThrSerAla.....IleGl 75
 468 TATCTCTCTCTCGCCCTAATAAAAAATAACACGCGCATGTGTTT 419
 :|||
 75 yglLeuAapLeuSerMetSerSerSerIleAlaThrArgAlaIleThr 92
 418 TGTATTTCTGTTTCTTATCATCTCTCTGTTGATATTTCTCTGTC 369
 |||||
 92 yrtYrPheValThrThrIleSerAlaValIleLeuGlyIleCysLeuVal 108

368 CCGTCTCTCCCCCCCCCATATATAATATCTGTTATTAATGG 319
 :|||
 109 ThrThrLeuArgProGlyGlnGlyAlaLysIleValGluThrGlnThrGl 125
 318 TGGGGTGATTAATTTTGGGGGAAAACTCATGTCCTCC 279
 :|||
 125 usrIleAspLys.....AlaSerLysValLeuThrPro 136
 seq_name: /SIDS8/gcdata/geneseq/AA2001.DAT:AA671446
 seq_documentation_block:
 ID AAG71446 standard; Protein: 247 AA.
 AC AAG71446:
 XX
 DT 31-JUL-2001 (first entry)
 XX
 DE Human olfactory receptor polypeptide, SEQ ID NO: 1127.
 XX
 DE Human olfactory receptor; OR: primary scent determination;
 KW Human olfactory receptor; OR: primary scent determination; polypeptide library; odour receptor;
 KW scent profile; scent fingerprint; scent representation.
 XX
 OS Homo sapiens.
 XX
 PN WO200127158-A2.
 PD 19-APR-2001.
 XX
 PP 06-OCT-2000; 2000WO-US27582.
 XX
 PR 08-OCT-1999; 99US-0158615.
 PR 24-FEB-2000; 2000US-0184809.
 XX
 PA (DIGI-) DIGISCENTS.
 PA (YEDA) YEDA RES & DEV CO LTD.
 XX
 PI Bellenson J, Smith D, Lancel D, Glusman G, Fuchs T, Yanai I;
 DR WPI: 2001-290713/30.
 XX
 PT New polynucleotides which encode polypeptides involved in olfactory
 PT sensation for identifying olfactory agonists and antagonists -
 XX
 PS Claim 11; Page 650-651; 1857pp; English.
 XX
 CC The present sequence is an olfactory receptor which is encoded by
 CC one of a number of novel polynucleotides. The polynucleotides can be
 CC used in screening for olfactory agonists and antagonists. The methods
 CC allow for the determination of primary scents and the identification
 CC of the odour receptors used to detect these primary scents. The methods
 CC also enable determination of secondary scents and the identification of
 CC combinations of odour receptors that are involved in detecting such
 CC secondary scents. This enables the construction of a scent representation
 CC (also called a scent fingerprint or scent profile), which may be used to
 CC re-create and edit scents. Libraries of olfactory receptors are useful
 CC for determining the interaction pattern of a composition with the
 CC receptors, and can be used for determining differences in the olfactory
 CC faculties of different individuals.
 XX
 SQ Sequence 247 AA:
 alignment_scores:
 Quality: 80.00 Length: 169
 Ratio: 0.976 Gaps: 8
 Percent Similarity: 48.521 Percent Identity: 23.077
 alignment_block:
 US-09-823-101-3/rev x AAG71446 ..
 Align seg 1/1 to: AAG71446 from: 1 to: 247

```

600 AATTTATTCCTGCTGTCGTCCTCCCTCC... CCCCCACACAAAAAT 554
    ||||| ..... |||||
10 Asn1YrPhe1leu1eu1gly1eu5er1asn**Pro1leu1Glu1Ile1Va 26
553 AGTTTGTTCCTTTAATTTTGTGTTTCTACTCTCTCCATTATTCCTCC 504
    ||||| ..... |||||
26 Ile1Phe1Val1leu1leu1Ile1Phe1cys1Phe1Met1r1leu1Ile1y1SL 43
503 TTTAAATAATTAATTAATTAATTTGTTTCAATTTTGAATTCCTCTGCC 454
    || ..... |||||
43 eu..... Phe1Ser1Ile1leu1Ser1Yr1leu..... 51
453 TATATAAAAAAATAATACACGCGCATGTGTTTTTGTATTCCTGTTTC 404
    ..... ||| ||| ..... |||||
52 ..... Asp1Ser1His1Pro1His1Thr... Pro1Arg1Yr1leu1Phe1 63
403 GTTTATTCATCTTCCTGTGTGATATTTTCCTGTCGCTCTCTCCCCC 354
    ||||| .....
63 r1Phe1leu1asp1rhe..... 67
353 CCCCCCTCATTAATAATATCT..... CTGTTTATTAATTTGGTGG 316
    ||| |||||
68 ..... Cys1Yr1Thr1Ile1Ser1Ile1Phe**Leu1Gln1Yr1asn1leu1Trp 82
315 GGTGATTAATTTTGGGGGAAAAAACTCATGTCTCCCCCATGCGGGATAGC 266
    || ..... ||| ..... |||
83 Gly..... Pro1Gln1Yr1asn1Ile1Se 89
265 TCCNCACANGTCACGTTGTCTGTATPACTCTGTGTTCAACCAANTGGACACA 216
    ..... ||| ..... |||
89 r1Yr1Ala1Ser1Met1Ile1Gln1Ile1Yr1Phe1Val1leu1Thr1leu1Gly1rhm 106
215 CAAGTTGGCTCTCATGTTGGTTGACCCACGTCATCGTGAACGTGATG... 169
106 et1As1Cys1Ala1leu1eu1Val1Met1Ser1Arg1r1Val1Yr1Ala1a1Gly 122
168 ... CCACAGTTTCCACGCACTGCTGTGCTGTCATGATATGACGTGTTGA 122
    ||||| ..... ||| ..... ||| ..... |||
123 His1Arg1His1leu1Pro1Yr1Thr1Val1Val1Met1Ala1Val1Ala1Phe1r1Pval1Se 139
121 GAATTAT 115
139 r1Ser1Phe 141

seq_name: /SIDS8/gcgdata/geneseq/geneseqp/AA2001.DAT:AA672416
seq_documentation_block:
ID -AA672416 Standard; Protein; 247 AA.
XX
XX AAG72416:
AC
XX
DT 31-JUL-2001 (first entry)
XX
XX Human OR-like polypeptide query sequence, SEQ ID NO: 2097.
DE
XX Human olfactory receptor; OR; primary scent determination;
XX secondary scent determination; polypeptide library; odour receptor;
XX scent profile; scent fingerprint; scent representation.
OS Homo sapiens.
MO200127158-A2.
PD 19-APR-2001.
XX
XX 06-OCT-2000; 2000MO-US27582.
XX
XX 08-OCT-1999; 99US-0158615.
XX
XX 24-FEB-2000; 2000US-0184809.
XX
XX (DIGT-) DIGISCENTS.
XX (YEDA ) YEDA RES & DEV CO LTD.
XX

```

[illegible]

```

106 eCAspCysAlaLeuValMetSerArgThrValIleAlaGly 122
168 ...CCACAGCTTCACGACGACCTGCTGTCATGATGATGCTTGA 122
123 HisArgHisLeuProTyrThrValValMetAlaValAlaPheTyrVal 139
121 GAATTAT 115
139 rSerPhe 141

seq_name: /SIDS8/gcdata/geneseq/geneseqp/AA1996.DAT:AA91016
seq_documentation_block:
ID   AA91016 standard; Protein: 607 AA.
AC   AA91016;
DT   06-SEP-1996 (first entry)
DE   PRSVPGLuc fusion protein product.
KW   Fusion protein; antibody; sea firefly luciferase; IgG recognition domain;
KW   Staphylococcus aureus; protein A; Streptococcus; protein G; primer; PCR;
KW   polymerase chain reaction; amplification; chemiluminescent marker;
KW   immunoassay.
OS   Synthetic.
FH   Key
FT   Peptide
FT   Protein
FT   Protein
XX   JF08027200-A.
XX   30-JAN-1996.
XX   18-JUL-1994; 94JP-0165223.
XX   18-JUL-1994; 94JP-0165223.
XX   18-JUL-1994; 94JP-0165223.
XX   (TORA ) TORAY IND INC.
XX   WPI; 1996-136341/14.
XX   Protein comprising luciferase fused to IgG binding domain - useful
XX   as diagnostic reagent in chemiluminescent immunoassays
XX   Example 7; Page 10-12; 16pp; Japanese.
XX   Novel fusion proteins (AA91016-9) comprising a region of a protein
XX   which recognises an IgG antibody linked to the N-terminal sequence of the
XX   sea firefly luciferase. The IgG recognition domain is pref. from the
XX   Staphylococcus aureus protein A or the Streptococcal protein G.
XX   Plasmids expressing the fusion genes were constructed using the primers
XX   AA13015-20. The fusion protein sequence given here is the product from
XX   the plasmid PRSVPGLuc (sequence not given in the specification) which
XX   contains the RSV promoter, a sequence coding for the mouse
XX   immunoglobulin (Ig) H chain V(NF) gene signal peptide and N-terminal 8
XX   amino acids, the protein G antibody binding Fc domain, mouse Ig H chain
XX   V(NP) gene CH2-CH3 intron, a one amino acid (glycine) spacer and the
XX   luciferase coding sequence. The novel fusion proteins can be used as
XX   chemiluminescent markers in immunoassays.
XX   Sequence 607 AA;

```

```

alignment_scores:
Quality: 78.50      Length: 91
Ratio: 2.066      Gaps: 4
Percent Similarity: 41.759      Percent Identity: 30.769

```

```

alignment_block:
us-09-823-101-3/rev x AA91016
Align seg 1/1 to: AA91016 from: 1 to: 607
255 TCACGTTGCTATA.....CTGCTTC 233
:::|||||
101 ThrArgCysAlaThrCysLysArgAspIleLeuSerAspGlyLeucCysG 117
232 AACCCANTGTGGACACACAAGTTGCTCTCATGTTGTTGACCAAGTCA 183
117 uAsnLysProGlyLysThrCysCys..... 125
182 TCGTGAACGTGATGCCACAGTTCCACGCACTTGCTTGTGTCATGCAT 133
126 .....ArgMetCysGlnTyrValIleGlu 133
132 ATGACTGTTGAGAATTATAGATACGCTAGACACAGATGATACGACAG 83
134 CysArgValGlu.AlaAlaGlyTyrPheArgThrPheTyrGlyLysArgP 150
82 ACTCACAAGAAGATCCAGGCAAGCACACACTGTGTGACAGCATAGGGTCC 33
150 hAsnPheGlnGluProGlyLysTyrValIleuAlaArgGly..... 163
32 ACAGCAGGAACCCAGTGCACA 12
164 ThrLysGlyGlyAspTyrPser 170

```

```

1  TITLE OF INVENTION: PROTEIN, NUCLEIC ACIDS ENCODING THE SAME AND METHODS OF
2  TITLE OF INVENTION: USE
3  NUMBER OF SEQUENCES: 22
4  CORRESPONDENCE ADDRESS:
5  ADDRESSEE: David A. Jackson, Esq.
6  STREET: 411 Hackensack Ave, Continental Plaza, 4th
7  STREET: Floor
8  CITY: Hackensack
9  STATE: New Jersey
10 COUNTRY: USA
11 ZIP: 07601
12
13 COMPUTER READABLE FORM:
14 MEDIUM TYPE: Floppy disk
15 COMPUTER: IBM PC compatible
16 OPERATING SYSTEM: PC-DOS/MS-DOS
17 SOFTWARE: Patentin Release #1.0, Version #1.30
18 CURRENT APPLICATION DATA:
19 FILING DATE: April 9, 1998
20 CLASSIFICATION: 800
21 ATTORNEY/AGENT INFORMATION:
22 NAME: Jackson Esq., David A.
23 REGISTRATION NUMBER: 26,742
24 REFERENCE/DOCKET NUMBER: 1340-1-013N
25 TELECOMMUNICATION INFORMATION:
26 TELEPHONE: 201-487-5800
27 TELEFAX: 201-343-1684
28 INFORMATION FOR SEQ. ID NO: 4:
29 SEQUENCE CHARACTERISTICS:
30 LENGTH: 456 amino acids
31 TYPE: amino acid
32 STRANDEDNESS: single
33 TOPOLOGY: linear
34 MOLECULE TYPE: protein
35 DESCRIPTION: HENT1
36 HYPOTHEICAL: NO
37 FRAGMENT TYPE:
38 ORIGINAL SOURCE:
39 ORGANISM: Homo sapiens
40
41 US-09-058-389A-4
42
43
44 alignment_scores:
45     Quality: 73.50          Length: 235
46     Ratio: 0.758           Gaps: 14
47     Percent Similarity: 41.277   Percent Identity: 22.553
48
49 alignment_block:
50     US-09-823-101-3/rev x US-09-058-389A-4 ..
51
52 Align seg 1/1 to: US-09-058-389A-4 from: 1 to: 456
53
54 648 ACTAATAATTTCTCTCGGCGCTCCCTCATGATGATTAATTAATTA 599
55      |||:::|||||:::|||||:::
56 36 ThrGlnTyrPheThrAsnArgLeuAspMet.....SerGlnAs 48
57
58 598 TTATCTCTCTC.....GTCT 582
59      |||:::|||||
60 48 nvaIserLeuvalThraIaGluLeuSerLysaspAlaGlnAlaSerAla 65
61
62 581 CTCCTCTCTCCCC...CCACACAAAAAATAGTTTTTGTCTTAAATT 535
63      ::|:::|||||:::|||||:::
64 65 IaprolaIaIaProleuProGluArgGlnSerLeuSerAlaIlePheasn 81
65
66 534 TTGTTTCTACTCTCTCTTATTCGCCCCCTTTAAAAATAAATACTCT 485
67      |||||:::|||||
68 82 AsnvalMetThrLeuLysAlaMetLeuProLeu.....Le 93
69
70 484 CTCTGTTTCATATTGTAATCTCTCTCTGCTAATAAAAAAATAAAC 435
71      |||:::|||||:::
72 93 uIeuThrIleTyrIleu.....AsuSerP 101
73
74 434 ACGGCGCAT..... 427

```

```

101 heluHISglNargIleProGlnSerValAlaArgIleLeuGlySerLeuVal 117
126 ..GTGTTTGTGATTTGTTGTTTCGTTTATTCATCTTCCTGTGTGATA 380
118 AlaIleLeuValIlePheLeuIleThrAlaIle..... 128
379 TTTCCTGTGTCCTGCTCTCCCTCC...CCCTCATATAAATATATC 333
129 ...LeuValIleValGlnLeuAlaPheLeuProPhePheValIleThrM 144
332 TTGTT.....ATTAATTGGTGGGT.....GGATTA 307
144 etIleIleValIleValLeuIleAsnSerPheGlyAlaIleLeuGlnGlySer 160
306 ATTTTGGGGGAAAAAACAATCATGTCCTCCCATC..... 274
161 LeuPheGlyLeuAlaGlyLeuLeuProAlaSerIleThrAlaProIleMe 177
273 .....GGGATAGCTCCNCAACAGTCAGTGTGCTATGCTACTGTT 234
177 tSerGlyGlnGlyLeuAlaGlyPhePheAlaSerValAlaIleMetIleCys 194
233 CAACACANTGTGGACAC..... 217
194 IaIleAlaSerGlySerGlnLeuSerGlnSerAlaIlePheGlyTyrPheIle 210
216 ACAAGTTGCTGTCATGTTGTCGACACGTCATCGTCACTGATCC 167
211 ThrIleCysAlaValIleIle.....LeuThrIleIleCysTyrLe 224
166 ACAGCTTCACGCACTGCTGTGTCATGATGATATGACTGTGAGATT 117
224 uGlyLeuProArg.....LeuGlnPheT 232
116 ATAGG 112
232 yArg 233
seq_name: /cgn2_6/ptodata/2/1aa/6a_COMB.pep:us-08-665-259-25
seq_documentation_block:
; Sequence 25, Application US/08665259
; Patent No. 6028173
; GENERAL INFORMATION:
; APPLICANT: Landes, Gregory M.
; APPLICANT: Burn, Timothy C.
; APPLICANT: Connors, Timothy D.
; APPLICANT: Dackowski, William R.
; APPLICANT: Van Raay, Terence J.
; APPLICANT: Klinger, Katherine W.
; TITLE OF INVENTION: NOVEL HUMAN CHROMOSOME 16 GENES,
; TITLE OF INVENTION: COMPOSITIONS, METHODS OF MAKING AND USING SAME
; NUMBER OF SEQUENCES: 73
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENZYME CORPORATION
; STREET: One Mountain Road
; CITY: Framingham
; STATE: Massachusetts
; COUNTRY: United States of America
; ZIP: 01701
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/665,259
; FILING DATE: 17-JUN-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Dugan, Deborah A.
; REGISTRATION NUMBER: 37,315

```

```

; REFERENCE/DOCKET NUMBER: IGS-9.1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (508) 872-8400
; TELEFAX: (508) 872-5415
; INFORMATION FOR SEQ. ID NO: 25:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1684 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-665-259-25

alignment_scores:
Quality: 72.00 Length: 116
Ratio: 1.263 Gaps: 7
Percent Similarity: 49.138 Percent Identity: 28.448

alignment_block:
US-09-823-101-3/rev x US-08-665-259-25 ..
Align seg 1/1 to: US-08-665-259-25 from: 1 to: 1684

690 TACTTGAGATTATTTTGTTCATATATAGTCGCCCCCACTAATA 641
289 PheLeuLeuPhePheLeuPheLeu..... 296
640 TTTCTCTGGGGCTCCCTCTCATAGTCGATTAATCTATAAT.....T 597
297 .....LeuIleAlaIleSerPheMetThrLeuLeuP 307
596 ATTCCTGCTGCTGTCCTCCCTCC.....TCCCCCCC 565
307 hecysValIleValIleProAsnValAlaValLeuSerArgSerAspPro 323
564 ACCAAAAAATAGTTTGTTCCTTAATTTTGTTCCTCTCTCTT 515
324 SerLeuValIleAlaPheLeuLeuLeuLeuIleSerThrIleSer.. 339
514 ATTCCTCCCTCTTAATAAATAATTAATCTCTCTCTTTCATATTTGTAAT 465
340 .PheSerPheMet.....ValSerThrPhePheSer..... 349
464 CTCTCTCTGCTTAATAAATAAATAACACGCGCATGTTTGTGTA 415
350 .....LysAlaAsnMetAlaAlaAlaPheGly 359
414 TTTCGTTTTCGTTTATCCATCTCTGTTGATATTTCTCTGTCCTC 367
360 PheLeuTyrPhePheThrTyrIleProTyr...PhePheValAlaPro 374

seq_name: /cgn2_6/ptodata/2/1aa/6a_COMB.pep:us-08-762-500-25
seq_documentation_block:
; Sequence 25, Application US/08762500
; Patent No. 6030806
; GENERAL INFORMATION:
; APPLICANT: Landes, Gregory M.
; APPLICANT: Burn, Timothy C.
; APPLICANT: Connors, Timothy D.
; APPLICANT: Dackowski, William R.
; APPLICANT: Van Raay, Terence J.
; APPLICANT: Klinger, Katherine W.
; TITLE OF INVENTION: NOVEL HUMAN CHROMOSOME 16 GENES,
; TITLE OF INVENTION: COMPOSITIONS, METHODS OF MAKING AND USING SAME
; NUMBER OF SEQUENCES: 83
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENZYME CORPORATION
; STREET: One Mountain Road
; CITY: Framingham
; STATE: Massachusetts
; COUNTRY: United States of America
; ZIP: 01701

```

```

1      COMPUTER READABLE FORM:
2      MEDIUM TYPE: Floppy disk
3      COMPUTER: IBM PC compatible
4      OPERATING SYSTEM: PC-DOS/MS-DOS
5      SOFTWARE: PatentIn Release #1.0, v
6      CURRENT APPLICATION DATA:
7      APPLICATION NUMBER: US/08/762,500
8      FILING DATE: 09-DEC-1996
9      CLASSIFICATION: 435
10     PRIOR APPLICATION DATA:
11     APPLICATION NUMBER: US 08/665,259
12     FILING DATE: 17-JUN-1996
13     PRIOR APPLICATION DATA:
14     APPLICATION NUMBER: PCT/US96/10466
15     FILING DATE: 17-JUN-1996
16     ATTORNEY/AGENT INFORMATION:
17     NAME: Dugan, Deborah A.
18     REGISTRATION NUMBER: 37,315
19     REFERENCE/DOCKET NUMBER: ICS-9.3
20     TELECOMMUNICATION INFORMATION:
21     TELEPHONE: (508) 872-8400
22     TELEFAX: (508) 872-5415
23     INFORMATION FOR SEQ ID NO: 25:
24     SEQUENCE CHARACTERISTICS:
25     LENGTH: 1684 amino acids
26     TYPE: amino acid
27     TOPOLOGY: linear
28     MOLECULE TYPE: protein
29     US-08-762,500-25

```

alignment_scores:	
Quality:	72.00
Ratio:	1.263
Percent Similarity:	49.138
	Percent Identity: 28.448
	Length: 116
	Gaps: 7

```

alignment_block:
US-09-823-101-3/rev x US-08-762-500-25    ..
Align seg 1/1  to: US-08-762-500-25  from: 1  to: 1684

```

```

690 TATTAGATGATTTTGTTCCTTCATTAAGTAGGGCCCCCAGCTAATAA 641
    ::::: ::::::::::::::
289 PheLeuIrpPhePheLeu..... 296
640 TTTCCTGGCGTCTCCCTCATAGTCGATAACTATAAT.....T 597
297 ..... LeuIleAlaIleSerPheMetThrLeuIrp 307
596 ATTCTCTCGTGTCTTCCTCC.....TCCCCCCC 565
    :: :: ::::: ::::: ::::: :::::
307 hecysValIysValIysProAsnValAlaValIleuserArgSerAspPro 323
564 ACCAAAAAATAGTTTGTCTCTTAATTTTGTTCCTCTCCCTT 515
    ::::: :::::::::::::: ::::: ::::: :::::
324 SerLeuValIleuAlaPheLeuIleCysPheAlaIleSerThrIleSer... 339
514 ATTCCCCCCCCCTTAAAAATAAATTAATACCTCTCTTTCATATTTGTAAT 465
    ||| ::::: ::::: :::::
340 .PheSerPheMet.....ValSerThrPheSer..... 349
464 CTCCTCTGCGCTAATAAAAAAATAACACAGCGCATGTGTTTTGTA 415
350 ..... LysAlaAsnMetValAlaAlaPheGly 359
414 TTTCGTGTTTCGTTATCCATCTTCCTGTGATATTTCCTGTGCC 367
    ||||| ::::: ::::: ::::: :::::
360 PheLeuThrPhePheThrTrpIleSerProIrp...PheIleValAlaIrp 374

```

```
seq_name: /cgn2_6/ptodata/2/1aa/6A_COMB.pep:US-08-762-500-75
seq_documentation_block:
: Sequence 75, Application US/08762500
```

Patent No. 6030806
GENERAL INFORMATION:
APPLICANT: Landes, Gregory M.
APPLICANT: Burn, Timothy C.
APPLICANT: Comors, Timothy D.
APPLICANT: Dackowski, William R.
APPLICANT: Van Raay, Terence J.
APPLICANT: Klinger, Katherine W.
TITLE OF INVENTION: NOVEL HUMAN CHROMOSOME 16 GENES,
TITLE OF INVENTION: COMPOSITIONS, METHODS OF MAKING AND USING SAME
NUMBER OF SEQUENCES: 83
CORRESPONDENCE ADDRESS:
ADDRESSEE: GENZYME CORPORATION
STREET: One Mountain Road
CITY: Framingham
STATE: Massachusetts
COUNTRY: United States of America
ZIP: 01701
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/762,500
FILING DATE: 09-DEC-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/665,259
FILING DATE: 17-JUN-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US96/10469
FILING DATE: 17-JUN-1996
ATTORNEY/AGENT INFORMATION:
NAME: Dugan, Deborah A.
REGISTRATION NUMBER: 37,315
REFERENCE/DOCKET NUMBER: IGS-9.3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (508) 872-8400
TELEFAX: (508) 872-5415
INFORMATION FOR SEQ ID NO: 75:
SEQUENCE CHARACTERISTICS:
LENGTH: 1704 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-762-500-75

[illegible]

```
alignment_block:
US-09-823-101-3/rev x US-08-762-500-75 .
```

```

Align seg 1/1 to: US-08-762-500-75 from: 1 to: 1704

690 TACCTGAGATTATTTTGTCTTCATTATAGTCGGCCCCACATAATA 6411
      ::::: ::::::::::::::
309 PheLeuLeuPhePheLeuPheLeu..... 316
640 TTCTCTCGGGCGCTCCCTCTCATAGCTGCTAATACTAAT.....T 597
      :::::::::::::: :::::
317 .....LeuIleAlaAlaSerPheMetThrLeuLeuP 327

596 ATTCTCTGCTGCTGCTCTCCCTCC.....TCCCCCCC 565
      ::: ::::: ::::: :::::
327 hecysVallylvallylproAlaAlaValLeuSerArgSerAsp 343
564 ACCAAAAAATATGTTTGTCTTATTTTGTGTTTCTACTCTGCTT 515
      ::::: ::::: ::::: :::::

```



```

; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 723 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Drosophila melanogaster
; IMMEDIATE SOURCE:
; CLONE: Drosophila SSRP (predicted)
; FEATURE:
; NAME/KEY: Domain
; LOCATION: 458..507
; OTHER INFORMATION: /label= Acidic
; FEATURE:
; NAME/KEY: Domain
; LOCATION: 518..547
; OTHER INFORMATION: /label= Basic I
; FEATURE:
; NAME/KEY: Domain
; LOCATION: 547..620
; OTHER INFORMATION: /label= HMG
; FEATURE:
; NAME/KEY: Domain
; LOCATION: 632..649
; OTHER INFORMATION: /label= Basic II
; FEATURE:
; NAME/KEY: Domain
; LOCATION: 657..723
; OTHER INFORMATION: /label= Mixed Charge
;
US-08-328-809-6
```

```

alignment_scores:
  Quality: 71.50      Length: 129
  Ratio: 1.212      Gaps:
  Percent Similarity: 45.736      Percent Identity: 23.256
```

```

alignment_block:
US-09-823-101-3 x US-08-328-809-6
```

```

Align seg 1/1 to: US-08-328-809-6 from: 1 to: 723
```

```

349 GGGGGGGGAGAGAGAGGAGACAGCAAGAAATATCAACAAGAGATGATG 398
|||||
511 G1yG1yG1yAspSerAspG1yAla1yLys1yG1yLys1yLysSerG1 527
399 TAAACGAAACAGCAATACAAAAACAC..... 426
527 uLy1yG1yLys1yG1yLys1ySH1ySG1uLyG1uArgThr1ySL 544
427 .....ATGCCGCTGCTATT 441
544 ysPProSerLysLysLysAspSerG1yLysProLysArgAlaThr 560
442 TTTTTTTAA.....TTAGCAGAGAGAGATTACAATATGAAAGAAGA 485
|||||
561 AlaPheMetLeuTrpLeuAsnAspThrArgG1uSer1yLeu1yAspG1uAs 577
486 G.....AGTAATTTATTTTAAAGGGGGGCAATAGAG 520
577 nProG1y1yLeu1yValThrG1u1yAla1yLysG1yG1yLysMetTrpL 594
521 AGAGTAGAACAACAAATTAAG..... 543
594 ysg1uLeuLysAspLysSerLysTrpG1uAspAla1yAla1yAspLys 610
544 .....AACAAAACTATTTTTGTGGGGGGGA.. 573
611 G1uArgTrp1yHisAspG1uMetArgAsnTrpLysProG1u1yAlaG1yLys 627
574 .....GGAGGAGACAGACAGAGAGA 594
|||||
```

```

627 pSerAspAsnG1uLysG1yLysSerSerLysLys 639
seq_name: /cgn2_6/ptodata/2/1aa/PCTrus_COMB.pep:PCT-US92-11107-11
seq_documentation_block:
; Sequence 11, Application PC/TUS9211107
; GENERAL INFORMATION:
; APPLICANT: Donahue, Brian A.
; APPLICANT: Toney, Jeffrey H.
; APPLICANT: Bruhn, Suzanne L.
; APPLICANT: Pil, Peter M.
; APPLICANT: Brown, Steven
; APPLICANT: Kelleet, Pauli
; APPLICANT: Essigmann, John M.
; APPLICANT: Lippard, Stephen J.
; TITLE OF INVENTION: DNA Structure Specific Recognition
; TITLE OF INVENTION: Protein and Uses Therefor
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: 2 Millia Drive
; CITY: Lexington
; STATE: MA
; COUNTRY: USA
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/11107
; FILING DATE: 19921218
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/539,906
; FILING DATE: 18-JUN-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: MIT-4787AAA
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-861-9540
; TELEFAX: 617-861-6240
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 723 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Drosophila melanogaster
; IMMEDIATE SOURCE:
; CLONE: Drosophila SSRP (predicted)
; FEATURE:
; NAME/KEY: Domain
; LOCATION: 458..507
; OTHER INFORMATION: /label= Acidic
; FEATURE:
; NAME/KEY: Domain
; LOCATION: 518..547
; OTHER INFORMATION: /label= Basic I
; FEATURE:
; NAME/KEY: Domain
; LOCATION: 547..620
; OTHER INFORMATION: /label= HMG-box
; FEATURE:
; NAME/KEY: Domain
; LOCATION: 632..649
; OTHER INFORMATION: /label= Basic II
; FEATURE:
; NAME/KEY: Domain
; LOCATION: 657..723
```

OTHER INFORMATION: /label= Mixed Charge
PCT-US92-11107-11

```
alignment_scores:
  Quality: 71.50
  Ratio: 1.212
  Gaps: 5
Percent Similarity: 45.736
Percent Identity: 23.256
```

```
alignment_block:
US-09-823-101-3 x PCT-US92-11107-11 .
```

Align seg 1/1 to: PCT-US92-11107-11 from: 1 to: 723

```

349 GGGGGGGGGAGAGACGGGACAGCAAGAAATATCAACAAGGAATGGA 398
|||||:|||||:|||||:|||||:|||||:
511 GtgcgtgclyspserseracscglalalyslyslyscltulsylsSercl 527
|||||:|||||:|||||:|||||:|||||:
399 TAAACGAAACAGAAATCAAAAAAC. 426
|||||:|||||:|||||:|||||:
527 ulslyscglulsylsgtulsylslyscltulscluarqthrlysl 544
|||||:|||||:|||||:|||||:
427 .....ATGGCGTGGTATT 441
|||||:|||||:
544 yspSerLysLysLysLysAspSerclYlYsProlYsArqAlaThrThr 560
|||||:|||||:|||||:|||||:
442 TTTTTTTT.....TTAGCAGAGAGATTTACAAATATCAAAAGAGA 485
|||||:|||||:|||||:|||||:
561 AlaphemclLeuThrLeuAsnAspThrArgclUsSerlLysArgclUs 577
|||||:|||||:|||||:|||||:
486 G.....AGTATTATTATTTTAAAGGGGGGGAATAGAG 520
|||||:|||||:|||||:|||||:
577 nProclYlLysValThrGlnLeaLysLysGtYlYsclUmetrPl 594
|||||:|||||:|||||:|||||:
521 AGAGTAGAACAACAAATTAAG..... 543
|||||:|||||:|||||:|||||:
594 ysglUleuLysAspLysSerlYsPrglUnAspAlaAlaLysAspLys 610
|||||:|||||:|||||:|||||:
544 .....AACAAAACATATTTTGGTGGGGGGA.. 573
|||||:|||||:|||||:|||||:
611 GlnArTYrHIsAspGlUmetArqAsnTYlYsPrgclUnlAclYlYs 627
|||||:|||||:|||||:|||||:
574 .....GGAGGAGACACGACGAGAGA 594
|||||:|||||:|||||:|||||:
627 pSerAspAsnclUlsYsclYlYsSerSerLysLys 639
|||||:|||||:|||||:|||||:
seq.name : /cgn2_6/plodata/2/laa/6A.COMB.pep:US-08-726-214-8

```

seq_documentation_block:

Sequence 8, Application US/08726214
Patent No. 6107076
GENERAL INFORMATION:
APPLICANT: Tang, Wei-Jen
TITLE OF INVENTION: SOLUBLE MAMMALIAN ADENYLYL CYCLASE
TITLE OF INVENTION: AND USES THEREFOR
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSEE: Arnold, White & Durkee
STREET: P.O. Box 4433
CITY: Houston
STATE: Texas
COUNTRY: United States of America
ZIP: 77210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/726,214
FILING DATE: Concurrently Herewith
CLASSIFICATION: 435

PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 60/005,498
 FILING DATE: 04-OCT-1995
 ATTORNEY/AGENT INFORMATION:
 NAME: Highlander, Steven L.
 REGISTRATION NUMBER: 37,642
 REFERENCE/DOCKET NUMBER: UTSD:450
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (512) 418-3000
 TELEFAX: (512) 474-7577
 INFORMATION FOR SEQ ID NO: 8:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1064 amino acids
 TYPE: amino acid
 STRANDEDNESS:
 TOPOLOGY: linear
 US-08-726-214-8

```
alignment_scores:
  Quality: 71.00
  Ratio: 0.887
  Length: 175
Percent Similarity: 45.714
Percent Identity: 25.143
```

```
alignment_block:
US-09-823-101-3/rev x US-08-726-214-8 .
```

Align seg 1/1 to: US-08-726-214-8 from: 1 to: 1064

[illegible]

COUNTRY: U.S.A.

NUMBER OF SEQUENCES

```

CORRESPONDENCE ADDRESS:
ADDRESS: Knobbe, Martens, Olson & Bear
STREET: 501 West Broadway
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92101-3505

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: Win95
SOFTWARE: Word

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/905,223
FILING DATE:
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Israelson, Ned A.
REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET NUMBER:
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 417:
SEQUENCE CHARACTERISTICS:
LENGTH: 92 amino acids
TYPE: AMINO ACID
TOPOLOGY: LINEAR
MOLECULE TYPE: PROTEIN
ORIGINAL SOURCE:
ORGANISM: Homo Sapiens
TISSUE TYPE: Brain
FEATURE:
NAME/KEY: sig-peptide
LOCATION: -41...-1
IDENTIFICATION METHOD: Von Heijne matrix
OTHER INFORMATION: score 4.7
OTHER INFORMATION: seq LFSCFPLSHKFG/KK
US-08-905-223-417

```

```

alignment_scores:
Quality: 68.50      Length: 91
Ratio: 1.712       Gaps: 3
Percent Similarity: 43.956   Percent Identity: 32.967

```

alignment_block:

US-09-823-101-3/rev x US-08-905-223-417 ..

Align seg 1/1 to: US-08-905-223-417 from: 1 to: 92

```

539 AATTTTGTCTTCTACTCTCTCTATATCCGCCCTTTAAAAATAATT 430
    ||| |||:||||| ||| |||||
    3 AsnAlaCysIleValIleuProIhProPro..... 14
489 ACTCTCTCTTCATATTTGTAATCTCTGCTGCTAATAAAAAAAA 440
    ||| ::||| ||| |||||
    15 ....SerLeuGlnProSerAlaSerLeuAlaProAsn..... 26
439 TAACCAAGCGCATGTGTTTGTATTCGTTTTCGTTATCCATCTTC 390
    |||: ||| ||| ||||| |||
    27 .....ArgPheLeuPheSerCysPheCysPheLeuSerHisLysPhe 40
389 CTGTGTGATTTTCCTGTCGCCGCTCTCTCTCCGCC.....CCCCATAT 343
    ::| ||| ||| |||||
    41 GILyLysIleValIleTyPheAsnTyLeuSerGluLeuHisGluHisLe 57
342 AAAATATCTCTTGTATTAATGTTGGGGGATTAATTTTGGGGGAAA 293
    :||| ||| |||: ||| |||
    57 uUgStyTyrAspGlnLeuValIleProGluValLeuArgTyrAspGluL 74
292 AACTCATGTCGCCCATCGGGGA 270
    |||: ||| |||: |||

```

74 yslEugInserLeuHisGluGly 81

seq_name: /sgn2_6/Plodata/2/1aa/5A.COMB.pep:US-07-672-483-4

```

seq_documentation_block:
Sequence 4, Application US/07672483
Patent No. 5359029
GENERAL INFORMATION:
APPLICANT: LACROIX, Martial
TITLE OF INVENTION: PEPTIDES AND ANALOGUES AND MIXTURES
TITLE OF INVENTION: THEREOF FOR DETECTING ANTIBODIES TO HTLV-I AND HTLV-II
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: FISH & NEAVE
STREET: 875 Third Avenue
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10022-6250
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/672,483
FILING DATE: 19910302
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/554,258
FILING DATE: 18-JUL-1990
ATTORNEY/AGENT INFORMATION:
NAME: HALEY Jr, James F
REGISTRATION NUMBER: 27,794
REFERENCE/DOCKET NUMBER: 1AF8 CIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212)715-0600
TELEFAX: (212)715-0674
TELEX: 14-8367
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 433 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-672-483-4

```

```

alignment_scores:
Quality: 68.50      Length: 110
Ratio: 1.457       Gaps: 6
Percent Similarity: 42.727   Percent Identity: 26.364

```

alignment_block:

US-09-823-101-3/rev x US-07-672-483-4 ..

Align seg 1/1 to: US-07-672-483-4 from: 1 to: 433

```

529 TTCTACTCTCTCTTATTCGCCCTTTAAAAATAAT.....TA 489
    ||| ::||| ::| ||||| |||
    49 PheLeuLysLeuAlaLeuLysThrProIleTyPheAsnProIleAspTy 65
488 CTCTCTCTCTTCATATTTGTAATCTCTGCTGCTAATAAAAAAAAAT 439
    ||| ||| ||| ||||| |||
    65 rSerLeuLeu.....AlaSerLeuIleProLysGlyT 76
438 AACCAAGCGCATGTGTTTGTATTCGTTTTCGTT..... 401
    ||| ||| |||: ||| |||
    76 yPProGluArgValValGluIleIleAsnIleLeuValLysAsnGluVal 92
400 .....TATCATGTCCTCTGTGATTAATTTC 375

```

93 SerProSerAlaProAlaAlaProValProThr.Pro.....I 105
374 TGTGTCCGCTCTCTCCGCCCATATAAATATGCTGTATT 325
105 IecysProThrThrProProProProProSer..... 118
324 AATGTGTGGGTGATTAATTTTGGGGGAAACATCATGTCCCATC 275
119ProGluAlaHisValProProPr 126
274 GGGGATAGCTCCNCACANGTCACGTTGT 247
126 oTyValGluProThrThrThrGlnGys 135

seq_name: /cgn2_6/prodata/2/1aa/5A_COMB.pep:US-08-118-270-71

seq_documentation_block:
; Sequence 71, Application US/08118270
; Patent No. 5508384
; GENERAL INFORMATION:
; APPLICANT: Murphy, Randall B.
; TITLE OF INVENTION: POLYPEPTIDES OF G-COUPLED PROTEIN
; NUMBER OF SEQUENCES: 348
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK
; STREET: 419 Seventh Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/118,270
; FILING DATE: 09-SEP-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/943,236
; FILING DATE: 10-SEP-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Townsend, Kevin G.
; REGISTRATION NUMBER: 34,033
; REFERENCE/DOCKET NUMBER: MURPHY-2A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; TELEX: 248633
; INFORMATION FOR SEQ ID NO: 71:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 349 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-118-270-71

alignment_scores:
Quality: 68.00 Length: 150
Ratio: 1.193 Gaps: 7
Percent Similarity: 38.000 Percent Identity: 22.000

alignment_block:

US-09-823-101-3/rev x US-08-118-270-71

Align seg 1/1 to: US-08-118-270-71 from: 1 to: 349

599 ATGTATCTCTGTCGTCTCTCCCTCCGCCCAACCAAAAAATAGTT 550

147 IleValArgSerLeuCysLeuAspProPheProGluProAlaAsp..... 161
549 TTGTCTCTTAATTT.....TGTCTCTCTCTCTCTCTATT 512
162 LeuPheTrpLysLysLysPheAlaThrPheLeuLeuLeuLeuL 178
511 CCCCCCCTTT.....AAAAATAATTAATCTCTCTCTTT 477
178 eutProLeuPheIleIleSerValAlaTrpAlaArgValAlaLysLysLeu 194
476 CATTTTGTAT.....CTCTCTCT 457
195 TrpLeuCysAsnThrIleGlyAspValThrThrGluGlnTrpLeuAlaLe 211
456 GCCTAATAAATAAATAACACGCGCATG..... 426
211 uArgArgLysLysLysThrThrValLysMetLeuValLeuValValL 228
425TGTTTT 420
228 eutPheAlaLeuCysTrpPheProLeuAsnGlyTrpValLeuLeuLeuSer 244
419TTGTATTTC..... 411
245 SerLysAlaIleHisThrAsnAlaLeuTyPheAlaPheHisTrpPh 261
410TGTTCGTTAT 398
261 eAlaMetSerSerThrCysTyrsnProPheIleTyrcysTrpLeuAsnG 278
397 CCATCTTCCTTGATATTCCGTGTCCTCCGCTCTCTCCGCCCT 348
278 IuAsnPheArgValGluLeuLysAlaLeuLeuSerMetGlnProProPro 294

OM of: US-09-823-101-3 to: PIR_68: * out_format : pfs
 Date: Jan 17, 2002 4:03 PM

About: Results were produced by the Gencore software, version 4.5,
 Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODE=lfirmer-n2p.model -DEV=xlp
 -O=cpn2.1/USPRO.spool/US09823101/runat.17012002.155137_22030/app_query.fasta.1.2239
 -DB=PIR_68 -QFMT=fastan -SUFFIX=n2p.rpr -GAPOP=12.000
 -GAPEN=4.000 -MINMATCH=0.100 -LOOPEXT=0.000
 -OGAPOP=4.500 -OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
 -DELPOP=6.000 -DELEXT=7.000 -SPART=1 -MATRIX=blomsurf62
 -TRANS=human4.cdi -LIST=45 -DOCALLIGN=200 -THR_SCORE=pct
 -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
 -NORM=ext -MINLEN=0 -MAXLEN=200000000
 -USER=US09823101.ecgn1_1_200 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPHY
 -WAIT -THREADS=1

Search information block:

Query: US-09-823-101-3
 Query length: 698
 Database: PIR_68: *
 Database sequences: 219241
 Database length: 76174552
 Search time (sec): 120.350000

score_list:

Sequence	Start	Orig	ZScore	Escore	Len	Documentation
PIR2:A33723	90.50	161.75	0.0919	555	1	Cryptidina-luciferin 2-monooxygenase
PIR2:T47488	86.00	155.73	0.2860	373	1	hypothetical protein F9K21.60
PIR2:T27265	85.00	153.21	0.3851	396	1	hypothetical protein Y62H9A.10
PIR2:H86373	82.00	143.09	0.8545	653	1	protein T23E23.16 [imported]
PIR2:H70108	82.00	141.75	0.8578	773	1	hypothetical protein B8072.2
PIR2:T02001	81.00	141.07	1.11	652	1	hypothetical protein F15B16.10
PIR2:T37929	79.50	141.74	1.63	408	1	probable major facilitator protein
PIR2:T44522	77.50	137.31	2.75	426	1	hypothetical protein 4S [imported]
PIR2:T21502	77.00	137.12	3.13	384	1	hypothetical protein F28D9.2
PIR2:T44511	75.50	133.32	4.24	421	1	hypothetical protein 3P [imported]
PIR2:A83133	75.00	132.34	5.29	419	1	probable MFS transporter PA4096
PIR2:E71551	75.00	130.82	5.31	507	1	probable monooxygenase - Cblamy
PIR2:C81704	74.50	129.82	6.06	506	1	monooxygenase-related protein T
PIR2:T14191	74.00	131.73	6.84	350	1	extensin homolog T28D5.60 - Arg
PIR2:B96724	74.00	125.55	6.97	760	1	hypothetical protein F20P5.14
PIR2:G84511	73.50	125.64	7.92	661	1	hypothetical protein At2g13900
PIR2:A49580	73.50	125.54	7.92	670	1	mediates transport of organic a
PIR2:S71227	73.00	131.36	8.85	284	1	extensin 1 - Arabidopsis thalia
PIR2:T21855	73.00	129.67	8.89	351	1	hypothetical protein F36D3.7
PIR2:S33069	73.00	127.92	8.94	437	1	vasoactive intestinal peptide r
PIR2:T25032	72.50	130.40	10.09	282	1	hypothetical protein T20D3.8
PIR2:T23067	72.50	129.34	10.12	322	1	hypothetical protein H12C20.5
PIR2:B96798	72.50	128.16	10.15	373	1	extensin [imported] - Arabidops
PIR2:T18597	72.50	119.75	10.40	1071	1	hypothetical protein AC3.5 - C
PIR2:W5LR1	72.00	134.05	11.34	157	1	E5 protein - rhesus papillomavi
PIR2:T49564	72.00	129.07	11.51	293	1	hypothetical protein B208.120
PIR2:T31779	72.00	128.10	11.54	331	1	hypothetical protein F38H12.2
PIR2:T06247	72.00	126.69	11.59	395	1	UNIFOLIATA protein (allele J18)
PIR2:S58064	72.00	122.73	11.72	649	1	hdc protein - fruit fly (Drosop
PIR2:T29220	72.00	122.27	11.73	687	1	hypothetical protein C08D8.2
PIR2:S51592	72.00	121.25	11.77	781	1	hypothetical protein C08D8.2
PIR2:JUN901	72.00	119.73	11.82	945	1	Xyln precursor - Ruminoecoccus
PIR2:A53188	72.00	115.03	11.98	1704	1	endopeptidase Clp ATP-binding c
PIR2:T31363	72.00	115.03	11.98	1704	1	ATP-binding cassette transport
PIR2:D86452	72.00	113.77	12.02	1994	1	probable ATP-binding cassette
PIR2:T34268	71.50	127.30	13.15	322	1	protein F6N18.13 [imported]
PIR2:S38011	71.50	121.35	13.37	679	1	hypothetical protein F46C8.7
PIR2:S33686	71.50	120.84	13.39	723	1	hypothetical protein YK1179c
PIR2:H64577	71.50	117.83	13.51	1055	1	hypothetical protein - fruit fl
PIR2:H72557	71.00	130.87	14.79	181	1	type I restriction enzyme R pr
						hypothetical protein Ape1748

PIR2:S50150 - 71.00 124.02 15.08 427 | gastric CCK-A receptor - rab
 PIR2:T20045 - 71.00 122.80 15.14 498 | hypothetical protein C49A1.5
 PIR2:I45832 - 71.00 119.91 15.26 715 | scinderin - bovine
 PIR2:T05374 + 71.00 119.43 15.28 760 | hypothetical protein F16G30.
 PIR2:A41542 - 71.00 116.74 15.40 1064 | adenylate cyclase (EC 4.6.1

seq_name: PIR2:A33723

seq_documentation_block:

Cryptidina-luciferin 2-monooxygenase (EC 1.13.12.6) precursor - ostracod (Vargula hilg
 N:Alternate names: crustacean luciferase
 C:Species: Vargula hilgendorffii (sea firefly)
 C:Date: 09-Mar-1990 #sequence_revision 09-Mar-1990 #text_change 26-May-2000
 C:Accession: A33723
 R:Thompson, E.M.; Nagata, S.; Tsuji, F.I.
 Proc. Natl. Acad. Sci. U.S.A. 86, 6567-6571, 1989
 A:Title: Cloning and expression of cDNA for the luciferase from the marine ostracod V
 A:Reference number: A33723; MUID:89367290
 A:Accession: A33723
 A>Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-555 <THO>
 A:Cross-references: GB:M25666; NID:q162546; PID:q162547
 C:Keywords: luminescence; monooxygenase; oxidoreductase

alignment_scores:

Quality: 90.50 Length: 134
 Ratio: 1.775 Gaps: 6
 Percent Similarity: 38.060 Percent Identity: 28.358

alignment_block:

US-09-823-101-3/rev x A33723 ..

Align seg 1/1 to: A33723 from: 1 to: 555

```

372 TGTCCCGTCTCTCCGCCCCCTCATATAAATATCTGTATTAA 323
|||||
22 CysProValGluAlaGluProProser..... 30
322 TTGTGGGGGTGATTAATTTTGGGGGAAACATATGCCCCCATCG 273
|||||
31 .....SerThrProThrValProThrSerC 39
272 GGATGCTCCAC.....ANGTCACGTTGGCTATA..... 241
|||||
39 ysgLAlaLysGluGlyLucySllLeasPThrArgCysAlaThrCysLys 55
240 .....CTCTGTTCAACCAANTGTGCACACACAAG 212
56 ArgaspLleLeuserAspGlyLeucGlyLysnLysProGlyLysThrCy 72
211 TTGTCTCTCATGTTGGTGGACACGTCATGTCATGTCATGTCACACG 162
|||||
72 scys..... 73
161 TTCACGACGCTTGTCTGTGCATGATGATGATGATGATGATGATGAT 112
|||||
74 .....ArgMetCysGlnIleValIleGluCysArgValGluAlaAlaGlu 88
111 ATACGCTGAGCAGATGATGATGATGATGATGATGATGATGATGATGAT 62
|||||
88 yTyrThrArgTrpPheTyrGlyArgPheAsnGlnGlnGlnGlnGlnGln 105
61 AAGCACCTGCTGACAGCATAGGCTGCACAGCAGCAGCAGCAGCAGCAG 12
|||||
105 ySTyrValIleuAlaLysGly.....ThrLysGlyGlyAspThrSer 118
seq_name: PIR2:T47488
seq_documentation_block:
hypothetical protein F9K21.60 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)

```

C>Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000
 C/Accession: T47488
 R/Jordan, N.; Bangert, S.; Wiedemann, R.; Voss, H.; Unsel, M.; Mewes, H.W.; Lemcke, K.
 submitted to the Protein Sequence Database, February 2000
 A/Reference number: Z24467
 A/Accession: T47488
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-373 <JOR>
 A/Cross-references: EMBL:AL138657
 A/Experimental source: cultivar Columbia; BAC clone F9K21
 C/Genetics:
 A/Map position: 3
 A/Introns: 92/3; 328/2; 353/3
 A/Note: F9K21.60

alignment_scores:
 Quality: 86.00 Length: 209
 Ratio: 0.977 Gaps: 10
 Percent Similarity: 42.105 Percent Identity: 22.967

alignment_block:
 US-09-823-101-3/rev x T47488 ..

Align seg 1/1 to: T47488 from: 1 to: 373

```

648 ACTAATATTTCTCTCGGCGTCCCTCATAGTCATTAATCTATA 599
      ||||| ||| ||||| ||||| ||||| ||||| |||||
70 ThrIysalValSerPheArgIleAsnHisIleSerMetPheAsp 86
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
598 T.....TATTCCTCGTGTGTCCTCCCTCCGCCACCAAAA 558
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
86 nProGluIlePheGluLeuValMetGlyArgSerValProLysAspLysL 103
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
557 AATAGTTTGTCTTTAAAT..... 537
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
103 yslIleAlaIleMet.AspGluValGlnArgIleArgIleGlnPhe 119
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
536 .....TTTGTCTTCTACTCTCTTATCCGCCCTTAAATA 494
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
119 rSerIleProPheLeuValAlaSerAsnGluIleLysPheValTyrL 136
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
493 AATTAATCTCTCTTTCATATTTGTAATCTCTCTCTTAATAAAA 444
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
136 yslLeuAlaIleGluThrLeuValSerAsnIleSerIleProArgProGln 152
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
443 AAATAACCGCGCATGTGTTT..... 420
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
153 LysLysThrCysGlyAsnCysPheAsnAspGlyIleLysGlyLysAsnMe 169
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
419 .....TTGATTTCTGTGTTTCGTTATTCATCT 392
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
169 tPheSerAlaAspLeuCysSerHisIleTyrPheCys..... 180
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
391 TCCTTGTGATATTTCTGTGTCCTGTCCTCTCCGCCCTCATATA 342
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
181 .....ValGluCysMetLysGlu.....HisIle 188
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
341 AATATCTCTCTTATTAATTGTTGGGGGTGATAATTTTGGGGGAAA 292
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
189 Glu.ValSer.....LeuAsnGluGlyLysLeuP 198
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
291 ACTGATCTCCCGCATGCGGATAGCTCCACACANGTCAGTTGTCTAT 242
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
198 roArgCysProHisAspGlyCysThrSerAsnLeu..... 209
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
241 ACTCTGTTCAACCAANTGTGACACACAAGTTGCTCATGTTGGTTG 192
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
210 .....ThrLeuArgSerCysAspHisLeuLeu 219
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
191 ACCACGTCATCGTGAACGTCATGCCACAGT.....TCCACGCACTTGC 148
      ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

219 rProLysGlnArgIleMetTrpLysArgIleLysGluLysLeuP 236
147 TTGTCGTGTCATCGATATGACTGT 125
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
236 roValCysAspArgPheHisCys 243
      ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

seq_name: p1r2:T27265

seq_documentation_block:

hypothetical protein Y62H9A.10 - Caenorhabditis elegans

C/Species: Caenorhabditis elegans

C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000

C/Accession: T27265

R/Steward, C.

submitted to the EMBL Data Library, October 1998

A/Reference number: Z20334

A/Accession: T27265

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-396 <WIL>

A/Cross-references: EMBL:AL032630; PIDN:CAA21566.1; GSPDB:GN00028; CESP:Y62H9A.10

A/Experimental source: clone Y62H9A

C/Genetics:

A/Gene: CESP:Y62H9A.10

A/Map position: X

A/Introns: 29/1; 85/1; 108/3; 145/1; 177/3; 211/3; 235/1; 272/3; 329/1; 385/1

alignment_scores:
 Quality: 85.00 Length: 219
 Ratio: 0.914 Gaps: 10
 Percent Similarity: 42.466 Percent Identity: 21.918

alignment_block:
 US-09-823-101-3/rev x T27265 ..

Align seg 1/1 to: T27265 from: 1 to: 396

```

677 TTTTGTCTTCTCATTTAGTCGGCCCCACCTAAATTTCTCGCGCT 628
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
138 TyrThrPhePheIleValIleThrSerPro.....Va 147
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
627 CTCCCTCTCATAGCTGCATTAATTAATTAATTAATTAATTAATTA 578
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
147 AlaValSerLeuLeuGlyTyrCysIleHisLeuAlaIleSerLeu 164
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
577 CTCCCTCCCCCACCACCAAAATAGTTTGTCTTAAATTTTGTGTT 528
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
164 euProLys.....PheValAspAsnIleCysPhe 174
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
527 CTACTCTCTCTTATTCCTCCCTTAAATAATTAATTAATTAATTA 478
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
175 AsnPheSerValLeuProTyrPheArgAsnTyrLeuAlaTyrGlnAr 191
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
477 TCATATTTGT..... 468
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
191 gmetLeuCysValPheIleProIleIleIleTyrPhePheIleTyrIle 208
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
467 .....AATCTCTCTCTG 456
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
208 rGmetLysLeuArgPheGlnArgTyrGlyAsnSerAsnSerValLeuAsn 224
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
455 CCTAATAAAAAAATAATTAACACAGCCAGTGTGTTTGTGATTTCTGTT 406
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
225 LysGlyMetArgHisLeuHisTyrValSerPheIleThrLeuAla.. 240
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
405 TCCTTATCATCTCTGTTGATATTTCTGTCCTGTCCTCTCTCTCC 356
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
241 .....SerValPheLeuValLeu..... 247
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
355 CCCCCCTCATTAATAATATCTTGTATTAATTAATTAATTAATTAAT 306
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
248 .....ProGluValTyrTrpTyrPheThrLeuPheGlyMetAspAla 262
      ||||| ||||| ||||| ||||| ||||| ||||| |||||

```



```

530 TTTCTACCTCTCTATTCCTCCCTTAAATAAATAATCTCTTC 481
    :      :      :      :      :      :      :      :
456 eSerValAlaIleTyrIleProIleMetLeuLysnTyrIscysL 473
480 TTTTCATATTTGTAAT.....CTCTCTCCCTCAAT 450
    ||      :      :      :      :      :      :
473 eSerPhePheValTyrAlaAlaLeuMetIleSerIleIleProLeu 489
449 AAAAAAATAAAC.....ACGCCAGTGTG 424
    :      :      :      :      :      :      :
490 AspleuGluIleThrLeuLysIleIleSerLeuLeuPheThrPheLeu 506
423 TTTTGTATTTCTGTTTGTATTCATCTCTGTTGATATTTCT 374
    |||||      :      :      :      :      :      :
506 lPhePheTyrPhe...PheThrLeuSerAlaPheLeuAlaasn..... 519
373 GTGTCGGCTCTCTCCCTCCCTCATATATAATCTCTTGTATTA 324
    ||      :      :      :      :      :      :
520 .....ProIleLeuIleSerPheMetPheLeuVal 529
323 ATT 321
    :      :
530 Leu 530
seq_name: p1r2:T02001

```

```

seq_documentation_block:
hypothetical protein T15B16.10 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 26-Feb-1999 #sequence_revision 26-Feb-1999 #text_change 24-Mar-1999
C:Accession: T02001
R:Stonking, T.; Smith, R.
submitted to the EMBL Data Library, November 1998
A:Description: The sequence of A. thaliana T15B16.
A:Reference number: Z14488
A:Accession: T02001
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-652 <STO>
A:Cross-references: EMBL:AF104919; NID:g3859590; PID:g3859608
A:Experimental source: cultivar Columbia
C:Genetics:
A:Map position: 4
A:Introns: 135/1
A:Note: T15B16.10

```

```

alignment_scores:
Quality: 81.00      Length: 209
Ratio: 0.931      Gaps: 17
Percent Similarity: 41.627      Percent Identity: 25.837

```

```

alignment_block:
US-09-823-101-3/rev x T02001 ..

```

```

Align seg 1/1 to: T02001 from: 1 to: 652

```

```

584 TGTCCTCCCTCTCCCTCCCTCCCTCA..... 561
    |||||      :      :      :      :      :
105 CysAlaLysTyrProProGluValIleAsnIleSerLysThrHisH 121
560 .....AAAAATAGTTTGTCTTATTTGTTTCTTCTTCTTCT 527
    ||      :      :      :      :      :
121 shiLysLeuAsnLeuLeuLysGluIleGlnPheAspTyrCysVal. 137
526 TACCTCTCTATTCCTCCCTCCCTTAAATAAATAATCTCTCT..... 483
    :      :      :      :      :      :
138 .....TrpHisProProGluValAsnHisThrLeuGluValAsn 150
482 ...TCPTTGTATTTGTAATCTCTCT.....CTGCC 454
    |||||      :      :      :      :      :
151 HisSerLysHisSerLeuHisProLeuLysLeuHisThrGluLeuP 167

```

```

453 TAAATAAAAAATAACACAGCCATGTGT..... 423
    |||||      :      :      :      :      :
167 oAsnTyrSerAspArgLysCysArgLeuCysAlaLysGluIleGluValG 184
422 ..TTTTGTATTTCTGTTTGTATTCATCTCTCTTGTGATATTTCT 375
    :      :      :      :      :      :
184 lLeuPheTyrHisCys...SerLeuCysasnPheThrLeuAspMetSer 199
374 TGTCCTCCCTCTCTCTCCCTCCCTCATATAAATAATCTCT..... 333
    |||||      :      :      :      :      :
200 CysVal.....LeuasnProProGln....ArgTyrLeuThrPsnle 212
332 .....TTGTATTAATTTGTTGGGTGGATTAATTTT 302
    |||||      :      :      :      :      :
212 ulysAlaHisAspHisGlnLeuThrLeuLeuProSerLeuArgSerPhe 229
301 TGGGGGAAAAATCATATCTCTCCCTCATGCGGAGATAGCTCCACAGTCA 252
    ||      :      :      :      :      :
229 eu.Cys...serAlaCysGlyLeuAsnGlyAspArgSerProTyrIleC 244
251 GTTG...TGCTATACCTCTGTTCAACCAACNTGTGACACACAGTTGTGCT 205
    :      :      :      :      :      :
244 sValGlnCysAspPheMetIle.....HisGlnValCysL 256
204 CTCATGTTGGTTGACACACATCATCTGTAACCTGCATGCCACAC..... 163
    |||||      :      :      :      :      :
256 euHisLeu.....ProArg.ValIleAsnIleAsnArgHisAsp 270
162 .GTTCACGACGT.....T 150
    |||||      :      :      :      :      :
270 gValSerArgThrSerValLeuGlyValAlaAsnSerValCysGlyVal 287
149 GCCTTGCTGTCATCGATATGACT 127
287 ysArgGlnLysValAspTyrThr 294
seq_name: p1r2:T37929

```

```

seq_documentation_block:
probable major facilitator protein - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 15-Sep-2000
C:Accession: T37929; T38038
R:McDougal, R.C.; Rajandream, M.A.; Barrell, B.G.; Bothe, G.; Pohl, T.
submitted to the EMBL Data Library, August 1997
A:Reference number: Z21763
A:Accession: T38038
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-166 <MUR>
A:Cross-references: EMBL:Z98598; PIDN:CA811245.1; GSPDB:GN00066; SPDB:SPAC1B3.19
A:Experimental source: strain 972h-; cosmid c1B3
C:Genetics:
A:Gene: SPDB:SPAC1B3.19
A:Map position: 1

```

```

alignment_scores:
Quality: 79.50      Length: 88
Ratio: 1.622      Gaps: 3
Percent Similarity: 55.682      Percent Identity: 25.000

```

```

alignment_block:
US-09-823-101-3/rev x T37929 ..

```

Align seg 1/1 to: T37929 from: 1 to: 408

```

588 TTATTCCTGCTGCTGCTCCCTCCCTCCACCAAAAAAATAGTTT 549
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
106 Leu1leAla1alysGln1alargVal1leuAsnPro1asGln1val1leuSer1h 122
548 TTGTTCTTAAATTTTGTTCCTCTCTCTCTCTCTCTCTCTCTCT 506
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
122 rCys1erPro1leuTrp1leSer1Val1lleTrp1leu1leuAsnPro1leuTrp 139
505 .....CCTTTAAATAATAATACCTCTCT 482
133 he1euProG1yl1eAla1Cys1Ser1Ala1asPme1l1e1leuAsnPhet1r1h 155
481 CTTTTCATATTTGTAATCTCCTCTCTG.....CTAATAAAAAAAA 441
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
156 Leu1met1h1le1Trp1h1eAla1Ser1Cys1G1Ser1Ty1Ala1lle..... 169
440 ATATACACGCGCATGTGTTTGTATTTCTGTTTCTGTTTATTCATCTT 391
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
170 .....Ty1Ala1Cys1Cys1Met1Ala1leu1h1r1Val1h1e1leuAsnPro1as 184
390 CCTTGTGATATTT 377
184 la1leu1leu1e1uPhe 188

```

seq_name: p1r2:T44522

seq_documentation_block:
 hypothetical protein 4S (Imported) - Shigella sonnei
 C:Species: Shigella sonnei
 C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jan-2000
 C:Accession: T44522
 R:Chida, T.; Okamura, N.; Yoshida, Y.; Ohtani, K.; Arakawa, E.; Watanabe, H.
 submitted to the EMBL Data Library, May 1999
 A:Description: Complete DNA sequence of the O-antigen (rtb) gene cluster in Plesiomonas
 A:Reference number: Z22787
 A:Accession: T44522
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-426 <CH1>
 A:Cross-references: EMBL:AB028135; PIDN:BA85073.1
 A:Experimental source: Isolate HM583

alignment_scores:

Quality:	77.50	Length:	190
Ratio:	0.891	Gaps:	7
Percent Similarity:	45.789	Percent Identity:	22.632

alignment_block:

US-09-823-101-3/rev x T44522 ..

Align seg 1/1 to: T44522 from: 1 to: 426

```

677 TTTTCTTCTTCAATAGTCGCGCCGCCACATAATTTCTCGGCGT 628
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
93 Phe1met1h1er1Ser1Ty1Val1Ser1asPser1le1Ser1e1h1leu1e1Va 109
627 CTCCTCTCATAGTCATATACTATAATTAATTTCTCTGCTGCTCTCC 578
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
109 l1y1r1le1Ser1Ser1Ala1Ty1r1Phe1Ala1Cys1le1Ser1Ser1G1y1r1Phe1Ser 125
577 CTCGCCCCCCCCCCACCAAAA.....AATAGTTT 549
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
126 Leu1e1uG1ln1Ala1Val1G1y1r1Phe1Arg1Cys1G1u1Leu1Ty1le1Asn1le1Ty 142
548 TTGTTCTTAAATTTTGT.....TTCTACTCTCTCTCTAATTC 511
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
142 r1Ser1h1r1le1le1Ty1r1le1G1y1Cys1Asn1leu1Phe1leu1Ser1e1h1le1G 159
510 CCCCCCTTTAAATAATAATTAATTAATTAATTAATTAATTAATTC 461

```

```

159 l1u1Pro1.....Leu1Ty1r1Y1Ser1..... 164
460 TCCTGCTAATAAAAAAATAATACACGCGCATGTGTTTGTATTTTC 411
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
165 .....Ala1lle1Se 167
410 TGTTTTCGTTATTCATCTTCCTGTTGATATTTCTGTCGCGCTCT 361
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
167 r1le1Phe1le1Ty1r1Ser1le1Ser1leu1Val1Phe1Ser1Ser1h1s1Cys1A 184
360 CTCGCCCCCCCTCATATAAA..... 339
184 sn1Val1Pro1Cys1Phe1h1s1le1Lys1Arg1Pro1Ser1le1Leu1Val1Ty1r1Lys1asp 200
338 .....TATCTCTGTTATTAAT 322
201 Phe1leuAsn1PAla1h1r1Pro1Phe1Ala1lle1leu1Val1leu1leuAsn1Val1le 217
321 TGTGGGGGTGATAAATTTTGGGGGAAACATCATGTCGCCCATCGGG 272
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
217 u1Ser1le1leAsn1Phe1le1leu1Ty1r1G1u1Ty1r1Phe1Ser1Ty1ra1n..... 232
271 GATFACCTCCNCACANGTCACGTTGTGCTACTCTGTTCAACACANTGTG 222
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
233 .....Ser1Val1Ala1lle1Ty1r1Gln1Val1Val1h1r1Arg1Val 243
221 GACACACAAGTTTGCTC 204
244 Asn1h1r1Gly1leu1le1r1le 249

```

seq_name: p1r2:T21502

seq_documentation_block:
 hypothetical protein F28D9.2 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
 C:Accession: T21502
 R:Dobson, R.
 submitted to the EMBL Data Library, November 1996
 A:Reference number: Z19431
 A:Accession: T21502
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-384 <N1L>
 A:Cross-references: EMBL:Z81518; PIDN:CAB04215.1; GSPDB:GN00019; CESP:F28D9.2
 A:Experimental source: clone F28D9
 A:Genetics:
 A:Gene: CESP:F28D9.2
 A:Map position: 1
 A:Introns: 56/3; 97/3; 127/3; 227/2; 293/1

alignment_scores:

Quality:	77.00	Length:	204
Ratio: <td>0.837 <td>Gaps: <td>9 </td></td></td>	0.837 <td>Gaps: <td>9 </td></td>	Gaps: <td>9 </td>	9
Percent Similarity: <td>45.098 <td>Percent Identity: <td>22.549 </td></td></td>	45.098 <td>Percent Identity: <td>22.549 </td></td>	Percent Identity: <td>22.549 </td>	22.549

alignment_block:

US-09-823-101-3/rev x T21502 ..

Align seg 1/1 to: T21502 from: 1 to: 384

```

524 CTCCTCCCTTATTCGCCGCC...TTTAAATAATAATTAATCTCT 483
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
7 l1e1Cys1Pro1Ty1r1G1y1Pro1Phe1Asn1Ty1r1Met1Lys1h1n1leu1Gln1leu1 23
482 .....TCCTTCATATTTGTAATC 464
23 ec1Ly1e1uPhe1Ser1leu1Pro1leuAsn1Phe1leu1Cys1Phe1Ty1r1Phe1Val1Trp 40
463 TCCTCTGCTAATAAAAAAATAATACACGCGCATGTGTTTGTAT 414
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

```

```

40 hefYrThrProLysGluSerLysPhe.....ArgTYrCysLeuAlaTyr 54
413 TTGCTGTTGGTTATGCCATCTCTGTTGATTTTCCTGCTGTC...CC 367
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
55 PheIlnPheIleAlaPheLeuValGluValAlaSPMetSerLeuValCyspr 71
366 GTCTCTCTCTCCCTCCCTCATATATAATCTCTGTTATTAATGCTG 317
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
71 OGlyPhe.....TyrLeuPheProLeuMetGlyG 81
316 GGGGGAATAATTTTGGGGAAAACTCATGTCCTCCCATCGGGGATAG 267
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
81 LyrAlaLeuAlaGluThrAsnArgLeuPheSerGlyHisGln..... 95
266 CTCNCACANGTCACGTTGCTACTCTGTTCAAAACANTGCGACAC 217
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
96 .....ThrValValPheGlyTyrPheLysPheSerPheGluLe 108
216 ACAAGTTGCTCTCATATGTTGTTGACACGTCATCGTGAACATGCA... 171
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
108 uProSerLeuLeuLeuCysPheIleTYrArgHisAsnAlaIalaAsnpr 125
171 ..... 171
125 heAsnProLysPheLysIleArgArgSerIleGlnTYrLeuValIleAla 141
170 ...TGGCACACGTTCCACGACCTGCTGT..... 144
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
142 ThrCysHisThrPheProPheValThrAlaIleCysLeuLeuLysSerGI 158
143 .CTGTCAATCATATGACTGTTGAGAAATTAAGATACGCTAGG..... 102
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
158 uLeuSerHisValGlnGlnValArgIleLeuGluLysValArgArgTYrL 175
101 .....ACCAGATGCATAGCAGAGACTCACAGAAGATCCAGGCCAAA 60
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
175 yslEuglnThrArgGluThrAlaAsnTYrThrGlnAsnTYrProAsnCys 191
59 GCACACCTGGG 48
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
192 LeuHisIleLeu 195

```

seq_name: p1r2:T44511

seq_documentation_block:

hypothetical protein 3p [imported] - Plesiomonas shigelloides
 C:Species: Plesiomonas shigelloides
 C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jan-2000
 C:Accession: T44511
 R:Chida, T.; Okamura, N.; Yoshida, Y.; Ohnani, K.; Arakawa, E.; Watanabe, H.
 submitted to the EMBL Data Library, April 1999
 A:Description: Complete DNA sequence of the O-antigen (rfb) gene cluster in Plesiomonas
 A:Reference number: Z22786
 A:Accession: T44511
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-421 <CHD>
 A:Cross-references: EMBL:AB025970; PIDN:BAA85008.1
 A:Experimental source: ATCC 14029

alignment_scores:

Quality:	75.50	Length:	190
Ratio:	0.868	Gaps:	7
Percent Similarity:	45.789	Percent Identity:	22.632

alignment_block:

US-09-823-101-3/rev x T44511 ..

Align seg 1/1 to: T44511 from: 1 to: 421

```

677 TTTTGTTCCTCATATAGTGGCCCTCAATAATATTCCTGCGGT 628
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

93 PheMetPheTYrSerTYrValSerAspSerIleSerLeuThrLeuLeuVal 109
627 CTCCTCTCTCATCTGCATATACATAATATTTCTCTCTGCTGCTCC 578
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
109 lTYrIleSer.SerAlaTYrPheValCysIleSerSerGlyArgPheSer 125
577 CTCCTCCCTCCCTCCACCAAAA.....AATAGTTT 549
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
126 LeuLeuGlnAlaValAlaGlyArgPheArgCysGluLeuTYrIleAsnIleTY 142
548 TTGTTCTTAATTTTGT.....TTCTACTCTCTCTTATTC 511
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
142 rSerThrIleIleTYrIleGlyCysAsnLeuPheLeuSerLeuPheIleG 159
510 CCCCCCTTTAAAAATAATTAATCTCTCTTTTCATATTGTAATCTCC 461
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
159 lupro.....LeuTYrTYrSer..... 164
460 TCCTGCTTAATAAAAAAATAAACACGCGCATGTGTTTGTATTC 411
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
165 .....AlaIleLe 167
410 TGTTCCTTTATTCATCTCTCTGTTGATATTTCTGCTGCTCT 361
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
167 rIlePheIleTYrSerIleSerLeuValPheSerSerHisLysCysAl 184
360 CTCCTCCCTCCCT.CATATAAA..... 339
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
184 snValProCysPheHisIleLysArgProSerLeuValTYrLysAsp 200
338 .....TACTCTGTTAATAAT 322
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
201 PheLeuAspAlaThrProPheAlaIleLeuValLeuLeuAsnValIle 217
321 TGTGGGGGTGATTAATTTTGGGGAAAAATCATATGTCCTCCCATCGGG 272
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
217 uSerSerIleAspLeuPheIleLeuLysGluTYrPheSerTYrAsn.... 232
271 GATAGCTCCACACANGTCAGTGTGCTATACCTGTCTCAACACANTGTC 222
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
233 .....SerValAlaIleTYrGlnValValThrArgVal 243
221 GACACACAGTTGTGCTC 204
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
244 AsnThrClyLeuMetIle 249

```

seq_name: p1r2:A83133

seq_documentation_block:

probable MFS transporter PA4096 [imported] - Pseudomonas aeruginosa (strain PA01)
 C:Species: Pseudomonas aeruginosa
 C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
 C:Accession: A83133
 R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.;
 Adam, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lapidis, K.; L
 ; Lory, S.; Olson, M.V.
 Nature 406, 959-964, 2000
 A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa
 A:Reference number: A82950; MUID:20437337
 A:Accession: A83133
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-419 <STO>
 A:Cross-references: GB:AE004826; GB:AE004091; NID:9950296; PIDN:AA607483.1; GSPDB:GN
 A:Experimental source: strain PA01
 A:Genetics:
 A:Gene: PA4096

alignment_scores:

Quality:	75.00	Length:	150
Ratio:	1.136 <td>Gaps:</td> <td>7 </td>	Gaps:	7
Percent Similarity:	44.000 <td>Percent Identity:</td> <td>25.333</td>	Percent Identity:	25.333

alignment_block:

US-09-823-101-3/rev x A83133 ..

Align seg 1/1 to: A83133 from: 1 to: 419

```

662 ATAGTCGGCCCCCACAATAATTTCTCGGCGTCCCTCTCATAGCT 613
      ::::::::::::::::::::
154 VALLIEGlyProProlLeuValAlaLeuAlaGluAlaLeuGlyTrpPar 170
      ::::::::::::::::::::
612 GCATAATCTAATATATCTCTGCTGCTC..... 579
      ||| |||::: ||| |||
170 gHisAlaPheIleValSerLeuProGlyLeuIleAlaTrpIleuI 187
      ::::::::::::::::::::
578 .....CCTCCTCCGCCCCACCAAAAAAATGATTG 546
      ||| |||::: ||| |||
187 leTrpArgHisValArgProAspProArgAlaGlnArg..... 200
      ::::::::::::::::::::
545 TTCCTTAATTTTGTCTACTCTCTCTTAATTCCTCCCTTAAAAA 496
      ::::::::::::::::::::
201 .....ProAlaProArgAlaGlnAlaGlySgl 210
      ::::::::::::::::::::
495 TAAATTAATCTCTCTTTTCATATTTGAATCTCCTCGCTAATAAA 446
      ||| |||::: ||| |||
210 yLysArgLeuAlaLeuLeuLysSerArgAsnIleLeu..... 223
      ::::::::::::::::::::
445 AAAAAATAACACGCGCATGTGTTTGTATTTCTGTTTCGTTATCC 396
      ||| |||::: ||| |||
224 .....CysThrIleuIleSerCysValPheValThr 233
      ::::::::::::::::::::
395 ATCTTCCTTTGATATTTCTGTGCTCCGCT..... 363
      ||| |||::: ||| |||
234 TrpPheIleValIleSerPheThrProThrPheLeuValAsnAlaAr 250
      ::::::::::::::::::::
362 ....CTCCTCCCGCCCTCATATATAATATCTCTGTTATTAATG 317
      ||| |||::: ||| |||
250 gGlyPheSerProAlaThrMetGlyArgLeuMetSerCysLeu. 266
      ::::::::::::::::::::
316 GCGTGATGAATTTTGGGGGAAAAACATCATGCCCCCATCGGGAT 269
      ::::::::::::::::::::
267 AlaTrpVal...ValTrpGly...PheAlaValProAlaIleSerAsp 280

```

seq_name: p1r2:E71551

seq_documentation_block:

probable monooxygenase - Chlamydia trachomatis (serotype D, strain UW3/Cx)
 C:Species: Chlamydia trachomatis
 C>Date: 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change 08-Oct-1999
 C:Accession: E71551
 R:Stephens, R.S.; Kaiman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitchell, Science 282, 754-759, 1998
 A:Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia trachomatis
 A:Reference number: A71570; MUID:99000809
 A:Accession: E71551
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-507 <ARN>
 A:Cross-references: GB:AE001288; GB:AE001273; NID:g3328545; PID:AMC67739.1; PID:g332854
 C:Experimental source: serotype D, strain UW-3/Cx
 C:Genetics:
 A:Gene: mmpA

alignment_scores:

Quality:	75.00	Length:	159
Ratio:	1.056	Gaps:	8
Percent Similarity:	44.654	Percent Identity:	27.673

alignment_block:

US-09-823-101-3/rev x E71551 ..

Align seg 1/1 to: E71551 from: 1 to: 507

```

545 TTCCTTAATTTTGTGTTT..... 528
      ||| |||::: ||| |||
215 PheValAsnPheValPheTyrAsnProTyrArgGlySerArgGlnLeu 231
      ::::::::::::::::::::
527 .....CTACTCTCTTAATTCCTCCCGCCCTTAAAAATAATTAC 488
      ::::::::::::::::::::
231 rLeuAlaAsnThrSerGlyProLeuSerSerLysPheLysAsnLysLeu 248
      ::::::::::::::::::::
487 TCCTCTCTTTTCATATTTGTAATCTCCTCGCTAATAAAAAATA 438
      ||| |||::: ||| |||
248 euTrpThrTyrGlyLeuAla....LeuAlaGluAspProLeuSerIle 262
      ::::::::::::::::::::
437 ACCACGCCCATGTGTTTTGTATTTCTGTTTCGTTATTCATCTCTCT 388
      ::::::::::::::::::::
263 SerSerSerLeuLeuGlnTyrProPheCysHisAspArgTyrIlePhe 279
      ::::::::::::::::::::
387 TGTGATATTTCCCTGTGTCGCCGTCTCTCTCCCGCCCTCATATAAAT 338
      ||| |||::: ||| |||
279 uGlySerIleAlaAsnAsnLeuSerPheSer..... 289
      ::::::::::::::::::::
337 ATCTCTGTATTAATTTGTTGGGGTG.....GATAAATTT 303
      ::::::::::::::::::::
290 .....TyrLeuSerGlyAlaAsnSerAsnIleHisAspAlaPhe 302
      ::::::::::::::::::::
302 ...TTGGGGAAAAATCATATGTCCTCCCGCATCGGGGATAGCTCCACANG 256
      ||| |||::: ||| |||
303 AsnLeuGlyTrpLysLeuLeu.ProValIleLysLysAlaAsnSerSerg 319
      ::::::::::::::::::::
255 TCACGTGTGCTATACCTGTTCAAACCAANTGTGCACACACAGTTGTC 206
      ::::::::::::::::::::
319 InLeuIleLeuSerLysGluLeuLys.....Thr 328
      ::::::::::::::::::::
205 TCTCATGTTGGTTGACACAGTCATGTAAGTCATGCGCACACGTTCCAC 156
      ||| |||::: ||| |||
329 SerHisVal...LeuProHisPheAsnGluValHisGlnLysAlaGln 344
      ::::::::::::::::::::
155 GCACCTGCTGTCTGTGCATCGATAT 131
      ||| |||::: ||| |||
344 LysLeuLeuPheSerAsnMetTyr 352

```

seq_name: p1r2:C81704

seq_documentation_block:

monooxygenase-related protein TC0425 [imported] - Chlamydia muridarum (strain N19g)
 C:Species: Chlamydia muridarum, Chlamydia trachomatis MOpn
 C>Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 11-May-2000
 C:Accession: C81704
 R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heldelberg, J.F.; White, O.; Hicke, C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg, Nucleic Acids Res. 28, 1397-1406, 2000
 A:Title: Genome sequences of Chlamydia trachomatis MOpn and Chlamydia pneumoniae AR39
 A:Reference number: AB1500; MUID:20150255
 A:Accession: C81704
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-506 <JET>
 A:Cross-references: GB:AE002309; GB:AE002160; NID:g7190464; PID:AAF39281.1; PID:g719
 C:Experimental source: strain N19g (MOpn)
 C:Genetics:
 A:Gene: TC0425

alignment_scores:

Quality:	74.50	Length:	156
Ratio:	0.993	Gaps:	8
Percent Similarity:	48.077	Percent Identity:	26.923

alignment_block:

US-09-823-101-3/rev x C81704 ..

Align seg 1/1 to: C81704 from: 1 to: 506

551 TTTTGTCTCTTAAT.....TTTGTGTTT 529

```

|||||:|||||:
218 PheValPheTyrAsnProTyrArgGlySerArgGlnLeuTyrLeuAlaAs 234
528 TCTACTCTCTCCCTTATTCCTCCCTTTAAATAATTTACTCTCTTCT 419
234 ntrSerGlyProLeuSerProLysPheLysAsnLysLeuLeuTyrThr 251
478 TTTCATATTTTAAATCTCTCTGCTATATAAAAAATAAACACGCGC 429
251 YrGlyLeuAla.....LeuAlaGlnAsnProLeuSerLieserAlaSer 265
428 ATGTGTTTTTGTATTTCTGTTTTCGTTTATTCACATCTCTGTTGAT 379
266 PheMetGlnTyrProPheGlySHisAsnArgTyrLephLeuGlySerVa 282
378 TTTCCTGTGTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTGT 329
282 LAlaAsnAsnLeuSerPheSer..... 289
328 TATTAATTTGCTGGGCTG.....GATAAATTT...TTGGGG 297
290 ..TyrLeuSerGlyValAlaAsnLeuAlaHisGlnAlaPheAsnLeuGly 305
296 GAAAAACATCATGTCCCTCCATCGGGGATAGCTCCNCAACAGTCAGTTGT 247
306 TrpLysLeuLeu..ProValLleLysLysAlaAlaSerSerGlnLeuLeL 322
246 GCATATCTCTGTCTCAACACANTGTGGACACACAAGTTGTCTCTCATGTT 197
322 euserArgGlnLeuLys.....ThSerHisVal 331
196 GGTGACCATCATCGTGAATGCATGCCACAGCTTCACGACCTTGCT 147
332 ..LeuProHisPheAsnGlnValHisGlnLysArgThrAlaLysLeuLe 347
146 TGTCTGTATCATGATAT 131
347 upHeserLysMetTyr 352
seq_name: p1r2:T14191

seq_documentation_block:
extensin homolog T28D5.60 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 24-Nov-1999
C:Accession: T14191
R:Bevan, M.; Lennard, N.; Quail, M.; Harris, B.; Rajandream, M.A.; Barrell, B.G.; Bancroft
submitted to the Protein Sequence Database, August 1999
A:Reference number: 217931
A:Accession: T14191
A:Molecule type: DNA
A:Residues: 1-350 <BEV>
A:Cross-references: EMBL:AL109819
A:Experimental source: cultivar Columbia; BAC clone T28D5
C:Genetics:
A:Gene: ATSP:T28D5.60
A:Map position: 4
A:Introns: 114/2; 184/2
C:Superfamily: hydroxyproline-rich glycoprotein

alignment_scores:
Quality: 74.00 Length: 122
Ratio: 1.370 Gaps: 7
Percent Similarity: 44.262 Percent Identity: 30.328

alignment_block:
US-09-823-101-3/rev x T14191
Align seg 1/1 to: T14191 from: 1 to: 350
672 TTTCATATTTATAGTCGCGCCGCCACTAAATATTTCTCTGCGGCTCTGCC 623
|||||:|||||:|||||:

```

```

241 PheIleTyrSerSerProProProProTyrValTyrLysSerValPr 257
622 TCTCATAGCTGCATATATATATATATATCTGCGTGGTCCCTCCT 573
257 oArgLlePro.....PheIleTyrSer..SerPro.....ProPro 268
572 CCCCCCACCACAAAAAATAGTTTTTGTCTTAAATTTTGTCTTACT 523
269 ..ProProTyrValTyrAsnSerLarProArgLleProPheIle..TyrS 284
522 CTCTCCTTATTCCTCCCTTTAAATAATTTACTCTCTCTTTTCATA 473
284 eSerLeuProProProProTyrValTyrAsnSerLarProArgLlePro 300
472 TTGTATATCTCTCTCTG.....CCTAAATAAAAAATAATACAC 433
301 PheIleTyrSerSerProProProProProTyrValTyrAsnSerAlaPr 317
432 GCGCATGTGTTTTTGTATTTCTGTTTTCGTTATTCATCTCTGTTG 383
317 oArgLleProPheIleTyr..... 323
382 ATATTTCTGTGTCCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 333
324 .....SerSerProProProHisHisMetPheThr 333
332 TTGTTATTAAT 321
334 SerLeuPheLeu 337
seq_name: p1r2:B96724

```

```

seq_documentation_block:
hypothetical protein F20P5.14 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: B96724
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,
ansen, N.F.; Hughes, B.; Hultz, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maitli, R.; Marzia
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: AB6141; MUID:21016719
A:Accession: B96724
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-760 <STO>
A:Cross-references: GB:AE005173; NID:92194126; PIDN:AB61101.1; GSPDB:GN00141
C:Genetics:
A:Gene: F20P5.14
A:Map position: 1

alignment_scores:
Quality: 74.00 Length: 186
Ratio: 1.057 Gaps: 10
Percent Similarity: 37.634 Percent Identity: 24.194

alignment_block:
US-09-823-101-3/rev x B96724
Align seg 1/1 to: B96724 from: 1 to: 760
680 TATTTTGTCTCTCATATAGTCGCGCC..... 651
|||||:|||||:|||||:
16 TyrPhe...PhePheIleValIleuProPheGlnSerLeuSerGlnPh 31
650 .....CCACTAAATATTTCTCTGCGGCTCTCTCTCTCATAGTCGCATA 608

```



```
31 eapSerProglInasnle.....G 38
    ||| |||||
607 ATACTAATATTCCTCGTCGTCCCTCCCTCCCCCACCACAAA 558
    ::::::::::: ||| |||||
38 luthrPhePheProIleSerSerIeuSerProValProProPro..... 52
557 AAATAGTTTTGTTCTTAATTTTGTTCCTACTCTCCCTTA..... 513
    ||||| |||
53 .....LeuProProSerSerAs 59
512 ....TCCCCCCCCCTTAAATAATTAATCTCTCTTTTCATATTTGTA 467
    ||||| |||
59 nProSerProProSerAsnAsnSerSerSer..... 70
466 ATCTCCTCCTCAATPAAAAAATAATACACCGCATGCTTTT... 420
    ::||| |||||
71 .....AspLysLysThrIleThrLysAlaValIleuIleThr 82
419 .....TTGATTTCTGTTTCGTTT 400
    ::||| |||
83 AlaAlaSerThrIleuValAlaGlyValPhePhePheCysIeuGlnAr 99
399 ATCCATCTTCCTT.....GTGATATT. 378
    |||: |||
99 gCysIleIleAlaArgArgArgArgValGlyProValArgValG 116
377 ..TCCTGTGCCGTCCTCTCCCGCCCT..... 348
    ::||| |||||
116 lUAsnThrIeuProProTyProProProMetThrSerAlaAlaVal 132
347 .....CATATAA 340
133 ThrThrThrThrIleuAlaArgGlyPheThrArgPheGlyVally 149
    ::|||
339 ATATCTCTTTTATTATTAATGCTGGGCTGATAATTTTGGGGGAAAAC 290
    |||: |||
149 sGlyIeuIleIeuAspGlnAsnGlyIeuAspValIeuTyTrpArgLysL 166
289 TCATGTCC 282
    |||: |||
166 eugInSer 168
```

THIS PAGE BLANK (USPTO)


```

-----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC
CC      EMBL; AL022727; CAA18783.1; -
CC      EMBL; AJ302547; CAC20472.1; -
CC      EMBL; AJ302548; CAC20473.1; -
CC      EMBL; AJ302549; CAC20474.1; -
CC      EMBL; AJ302550; CAC20475.1; -
CC      EMBL; AJ302551; CAC20476.1; -
CC      EMBL; AJ302552; CAC21440.1; -
CC      EMBL; AJ302553; CAC21441.1; -
CC      EMBL; AJ302554; CAC21442.1; -
CC      EMBL; AJ302555; CAC21443.1; -
CC      EMBL; AJ302556; CAC21444.1; -
CC      EMBL; AJ302557; CAC20477.1; -
CC      EMBL; AJ302558; CAC20478.1; -
CC      InterPro: IPR000276; GPCR_Rhodpsn.
CC      Pfam: PF00001; 7tm_1.1.
CC      PROSITE: PS00237; G-PROTEIN_RECIP_F1_1; 1.
CC      PROSITE: PS50262; G-PROTEIN_RECIP_F1_2; 1.
CC      KW      G-protein coupled receptor; Transmembrane; Glycoprotein;
CC      MultiGene family; Olfaction; Polymorphism.
CC      DOMAIN 1 28
CC      TRANSSEM 29 52 1 (POTENTIAL).
CC      DOMAIN 53 60 2 (POTENTIAL).
CC      TRANSSEM 61 82 2 (POTENTIAL).
CC      DOMAIN 83 103 3 (POTENTIAL).
CC      TRANSSEM 104 123 3 (POTENTIAL).
CC      DOMAIN 124 142 4 (POTENTIAL).
CC      TRANSSEM 143 161 4 (POTENTIAL).
CC      DOMAIN 162 198 5 (POTENTIAL).
CC      TRANSSEM 199 222 5 (POTENTIAL).
CC      DOMAIN 223 239 6 (POTENTIAL).
CC      TRANSSEM 240 262 6 (POTENTIAL).
CC      DOMAIN 263 275 7 (POTENTIAL).
CC      TRANSSEM 276 295 7 (POTENTIAL).
CC      DOMAIN 296 311 7 (POTENTIAL).
CC      DISULFID 100 192 8 (POTENTIAL).
CC      CARBOHYD 8 8 T -> A (IN 6M1-3*02).
CC      VARIANT 113 113 /FTID=VAR_010949.
CC      TRANSSEM 226 226 R -> Q (IN 6M1-3*02).
CC      VARIANT 228 228 /FTID=VAR_010950.
CC      VARIANT 261 261 V -> I (IN 6M1-3*03).
CC      VARIANT 261 261 /FTID=VAR_010951.
CC      SEQUENCE 311 AA; 34950 MW; 84201145B9D5A468 CRC64;
CC
CC      alignment_scores:
CC      Quality: 74.50 Length: 163
CC      Ratio: 0.968 Gaps: 10
CC      Percent Similarity: 47.239 Percent Identity: 23.926
CC
CC      alignment_block:
CC      US-09-823-101-3/rev x OZJ3_HUMAN ..
CC
CC      Align seg 1/1 to: OZJ3_HUMAN from: 1 to: 311
CC
CC      597 TATTCCTGCGTGTCTCCCTCCCTCC...CCCCCACCACAAAATAAGT 551
CC      ||| :::::||||: |||: ||| :::::
CC      14 TYRPhelleleValGlclyPheSerAsnIYrProHisLeuGluValValI 30
CC      ::|||::: :::::||||| :::::
CC      550 TTTTGCTCTTAATTTTGTCTTCTACCTGCTGTCCTATTGCCCCCCTTT 501
CC      ::|||::: :::::||||| :::::
CC      30 ePheValIValLeuIlePheIYrLeuMetIleuIleGlyAsnLeu 46

```

```

500 AAAAAATAATACCTCTCTTTTCATATTGTCAATCTCTCTCTGCTAA 451
      ::::::::::::::::::::|
47 .....PheIleIleIleuSerTyrLeu..... 54
450 TAAAAAAAATAACACGCGCATG.....TTTGTGATTCT 410
      ::::::::::| | | | |
55 .....ASPSeRhISleuHIStrIPromETyrPhenPheuSeRs 68
      ::::::::::| | | | |
409 GTTTCGTTATTCATCTCTCTGTGTGATATTTCCGTCCGCTCTC 360
      ::| | | | |
68 nleuSerPheleuasp.....LeucSTyrThrThrs 79
      ::| | | | |
359 TCCCCCCCCCTCATATAAATATCTCTGTATTGATGGGGGTGA 310
      ::| | | | |
79 eRSeRIlePro.....GlnleuValasInleuTyrpIgly... 90
      ::| | | | |
309 TAAATTTTGGGGGAAACATCATGTCGCCCATCGGGCATGCTCCNCA 260
      ::| | | | |
91 .....ProGluIySThrIleSerTyrAl 98
      ::| | | | |
259 CANGTCACGTTGCTATACCTCTGTTCACAAACAGTGGACACAGAT 210
      ::| | | | |
98 aelYcysmetIleGlnleuTyrPheValleuAlaleuGlyThrThrGlnC 115
      ::| | | | |
209 GTGCTCTCATGTGGTT.....GACCACGTCATCGTGAACGCAATG 169
      ::| | | | |
115 yValleuIleuValIleMetSerTyrAspArgTyrAlaIleValcysArg 131
      ::| | | | |
168 cCA...CACGT.....CCAGCAGACTTGC 148
      ::| | | | |
132 ProleuHIStrThrValleuMetHISProAlaGlyPheCys 144

seq_name: SwissProt_39:O8B8_HUMAN

seq_documentation_block:
ID O8B8_HUMAN STANDARD: PRT: 311 AA.
AC O15620;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE OLFACTORY RECEPTOR 8B8 (OLFACTORY RECEPTOR TPCR85).
GN OR8B8.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE OF 126-282 FROM N.A.
RC TISSUE=Testis;
RA MEDLINE=97224452; PubMed=9119360;
RX Vanderhaeghen P., Schumans S., Vassart G., Parmentier M.;
RT "Specific repertoire of olfactory receptor genes in the male germ
  cells of several mammalian species.";
RL Genomics 39:239-246(1997).
CC -!- FUNCTION: PUTATIVE ODORANT RECEPTOR.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as their content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/).
CC or send an email to license@isb-sib.ch.
CC -----
DR EMBL: X89675; CAA61822.1; -
DR InterPro: IPR000276; GPCR_Rhodopsn.
DR Pfam: PF00001; 7tm.1; 1
DR PROSITE: PS00237; G_PROTEIN_RECPT_F1_1; 1
DR PROSITE: PS50262; G_PROTEIN_RECPT_F1_2; 1
DR G-protein coupled receptor; Transmembrane TM

```

156 CGCACTTGGCTTGTC 142

alignment_block:

ment_block:

Align seg 1/1 to: ENTL_HUMAN from: 1 to: 455

```

648 ACTAAATATTTCTCTCGCGCTCCCTCTCATCTGCAATATACTATA 599
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
35 ThrGlnTyrPheThrAsnArgLeuAspMet.....SerGlnAs 47
598 TTATCTCTCTGTC.....GTGT 582
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
47 nValSerLeuValThrAlaGluLeuSerLysAspAlaGlnAlaSerAla 64
581 CTCCTCCCTCCCTCC...CCACCAAAAAAATGTTTGTCTTAATTT 535
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
64 LarpAlaAlaIleProLeuProGluArgAsnSerLeuSerAlaIlePheAsn 80
534 TTGTTTCTACTCTCTCTCTTATTCCTCCCTTTAAAAATTAATTA 485
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
81 AsnValMetThrLeuGlyAlaMetLeuProLeu.....Le 92
484 CTCTCTTTCATATTGTAAATCTCTCTCTGCTAATAAAAAAATAC 435
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
92 uLeuPheThrTyrLeu.....AsnSerP 100
434 ACGGCGAT..... 427
      |||:|||||:|||||:|||||:|||||:|||||:
100 heLeuHISGlnArgIleProGlnSerValArgIleLeuGlySerLeuVal 116
426 ...GTCTTTTGTATTTCTGTTTCTGTTATTCATCTCTCTGTGATA 380
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
117 AlalLeuLeuValPheLeuIleThrAlaIle..... 127
379 TTTCCTCTGTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 333
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
128 ...LeuValLysValGlnLeuAspAlaLeuProPheValIleThrM 143
332 TTGTT.....ATTAATGTGGGGT.....GGATAA 307
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
143 etIleLysIleValLeuIleLysSerPheGlyAlaIleLeuGlnGlySer 159
306 ATTTTGGGGGAAAAACTCATGTCCCTCCATTCG..... 274
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
160 LeuPheGlyLeuAlaGlyLeuLeuProAlaSerLysIleThrAlaProIle 176
273 .....GGGATAGCTCCNCACANGTCACGTTGGCTATACTGTT 234
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
176 tSerGlyGlnGlyLeuAlaGlyPhePheAlaSerValAlaMetIleCys 193
233 CAACCCANTGTGGACAC..... 217
      |||:|||||:|||||:|||||:|||||:|||||:
193 IalIleAlaSerGlySerGluLeuSerGluSerAlaPheGlyTyrPheIle 209
216 ACAAGTTGCTCTCATGTTGGTTGACACGTCATCGTGAACGACATGCC 167
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
210 ThrAlaCysAlaValIleIle.....LeuThrIleIleCysTyrLe 223
166 ACACGTTCCACGACCTTGTTCTGTGCATCGATATACGTTGAGAATT 117
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
223 uGlyLeuProArg.....LeuGluPheThr 231
116 ATAGG 112
231 YIAR 232
seq_name: SwissProt_39:OATP_RAT
seq_documentation_block:
ID OATP_RAT STANDARD: PRT: 670 AA.
AC P46720.
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE SODIUM-INDEPENDENT ORGANIC ANION TRANSPORTER (ORGANIC ANION
DE TRANSPORTING POLYPEPTIDE).
GN SLC21A3 OR OATP.

```

```

OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=Liver;
RX MEDLINE=94105118; PubMed=8278353;
RA Jaquemin E., Hagenbuch B., Stieger B., Wolkoef A.W., Meier P.J.;
RT "Expression cloning of a rat liver Na(+)-independent organic anion
RT transporter.";
RL Proc. Natl. Acad. Sci. U.S.A. 91:133-137(1994).
CC -|- FUNCTION: MEDIATES THE NA(+)-INDEPENDENT TRANSPORT OF ORGANIC
CC ANIONS SUCH AS BROMOSULFOBROMOPHTHALEIN (BSF) AND CONJUGATED
CC (TAUROCHOLATE) AND UNCONJUGATED (CHOLATE) BILE ACIDS.
CC -|- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (PROBABLY).
CC -|- TISSUE SPECIFICITY: A HIGH LEVEL EXPRESSION IS SEEN IN THE LIVER
CC AND KIDNEY, WHILE A LOWER LEVEL EXPRESSION IS SEEN IN THE BRAIN,
CC LUNG, SKELETAL MUSCLE AND PROXIMAL COLON.
CC -|- PTM: GLYCOSYLATED.
CC -|- SIMILARITY: BELONGS TO THE SLC21 FAMILY OF TRANSPORTERS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; LI9031; AAA16451.1; -
CC DR InterPro; IPR002350; Kazal.
CC DR SMART; SM00280; KAZAL.1.
CC KM Transmembrane; Transport; Glycoprotein.
CC FT TRANSMEM 21 38
CC FT TRANSMEM 86 106 POTENTIAL.
CC FT TRANSMEM 156 176 POTENTIAL.
CC FT TRANSMEM 194 214 POTENTIAL.
CC FT TRANSMEM 246 266 POTENTIAL.
CC FT TRANSMEM 313 333 POTENTIAL.
CC FT TRANSMEM 336 376 POTENTIAL.
CC FT TRANSMEM 388 408 POTENTIAL.
CC FT TRANSMEM 515 535 POTENTIAL.
CC FT TRANSMEM 571 571 POTENTIAL.
CC FT TRANSMEM 602 622 POTENTIAL.
CC FT CARBOHYD 62 62 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 124 124 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 135 135 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 492 492 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC SQ SEQUENCE 670 AA; 74178 MW; 3D53A4E8E1E21536 CRC64;

```

alignment_scores:

Quality: 73.50 Length: 144
 Ratio: 0.967 Gaps: 9
 Percent Similarity: 52.778 Percent Identity: 26.389

alignment_block:

US-09-823-101-3/rev x OATP_RAT ..

Align seg 1/1 to: OATP_RAT from: 1 to: 670

```

595 TTCTCTCGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 548
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
332 PheIleAsnLysPheThrPheLeuProLysTyrLeuGlnGlnTyrG 348
547 .....TGTTCTTAATTTTGTGTTTC.....TACTCTCTCTTA 514
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
348 LysSerThrAlaGlnAlaIlePheLeuIleGlyValTyrSerLeu... 363
513 TTCCCCCCCCCTTAAAAATAATTAATCTCTCTCTTTCATATTGTATC 464
      |||:|||||:|||||:|||||:|||||:|||||:|||||:
364 .....ProPheIleCysLeuGlyTyr...LeuIleGlyGlyPheIleMet 377

```



```

463 TCCTCTGCTGCTAATTAATAAATACACGCGCATGTTTGTAT 414
      ::::: ||| ||| ::::: |||
378 LysLysPheLysIleThrValLysLysAlaIleThrValLysAlaPheCysLe 394
      ::::: ||| |||
413 TTTCTTTTCTGTTATTCATCTCTCTGTGATTAATTCCTGTCCTCCGTC 364
      ||||| ||| |||
394 uSerValIleGluTyr.....LeuLeu.PheLeuCysHisPhe 406
      ||||| |||
363 TCTCTCCCCCCCCCTCATATAATATCTGTATTATTAATGCTGGG 314
      |||
407 MetLeuThrCysAspAsnAlaAlaIleGluLeuThrThrSeryTyr 423
      |||
313 TCGATAAATTTTGGGGAAACATCATGTCGCCCATCGGAGATAGCTC 264
      |||
423 sGly.....ValGlnHisGlnLeuHisValGluSerLysValIleAla 438
      |||
263 CACACANCTACAGTTGCTGCTACTCTGTCAAC..... 229
      ::::: ||| |||
438 sPCysAsnThrArgCysSer.....CysSerThrAsnThrTrpAspPro 452
      |||||
228 CANTGTGACACACAAAGTTGCTCTCATG 199
      |||||
453 ValCysGlyAspAsnGlyValAlaIleTyrMet 462

seq_name: SwissProt_39:VIPs_RAT

seq_documentation_block:
ID VIPs_RAT STANDARD; PRT; 437 AA.
AC P3500;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE VASOACTIVE INTESTINAL POLYPEPTIDE RECEPTOR 2 PRECURSOR (VIP-R-2)
DE (PITUITARY ADENYLATE CYCLASE ACTIVATING POLYPEPTIDE TYPE III RECEPTOR)
DE (PACAP TYPE III RECEPTOR) (PACAP-R-3).
DE VIPR2.
GN Rat
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=olfactory bulb;
RX MEDLINE=94039806; Pubmed=8224221;
RA Lutz E.-M., Sheward W.J., West K.M., Morrow J.A., Fink G.,
RA Harmer A.J.;
RT "The VIP2 receptor: molecular characterisation of a cDNA encoding a
RT novel receptor for vasoactive intestinal peptide.";
RL FEBS Lett. 334:3-8(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain cortex;
RX MEDLINE=95080140; Pubmed=7988457;
RA Ustin T.B., Bonner T.I., Mezey E.;
RT "Two receptors for vasoactive intestinal polypeptide with similar
RT specificity and complementary distributions.";
RL Endocrinology 135:2662-2680(1994)
RT
-1- FUNCTION: THIS IS A RECEPTOR FOR VIP AS WELL AS PACAP-38 AND -27,
-1- THE ACTIVITY OF THIS RECEPTOR IS MEDIATED BY G PROTEINS WHICH
-1- ACTIVATE ADENYLATE CYCLASE. CAN BE COUPLED TO PHOSPHOLIPASE C.
-1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
-1- SIMILARITY: BELONGS TO FAMILY 2 OF G-PROTEIN COUPLED RECEPTORS.
-----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z25885; CAA81104.1; -

```

```

DR EMBL; U09631; AAB60459.1; -.
DR PIR; S39069; S39069.
DR GCRdb; GCR_0794; -.
DR GCRdb; GCR_1034; -.
DR InterPro; IPR000832; GPCR_secretin.
DR InterPro; IPR001879; Hormn_rcptor.
DR Pfam; PF00002; 7tm_2; 1.
DR PRINTS; PR00249; GPCRSECRETIN.
DR PRINTS; PR00491; VASOACTIVEIPR.
DR PRINTS; PR01155; VIP2RECEPTOR.
DR SMART; SM00008; Hormr; 1.
DR PROSITE; PS00649; G_PROTEIN_RECEP_F2_1; 1.
DR PROSITE; PS00650; G_PROTEIN_RECEP_F2_2; 1.
DR PROSITE; PS0227; G_PROTEIN_RECEP_F2_3; 1.
KW G-protein coupled receptor; Transmembrane; Glycoprotein; Signal.
FT SIGNAL 1 437
FT CHAIN 23 437
FT DOMAIN 23 125
FT TRANSMEM 126 150
FT DOMAIN 151 157
FT TRANSMEM 158 177
FT DOMAIN 178 202
FT TRANSMEM 203 226
FT DOMAIN 227 239
FT TRANSMEM 240 261
FT DOMAIN 262 278
FT TRANSMEM 279 302
FT DOMAIN 303 327
FT TRANSMEM 328 347
FT DOMAIN 348 359
FT TRANSMEM 360 379
FT DOMAIN 380 437
FT CARBOHYD 57 57
FT CARBOHYD 87 87
FT CARBOHYD 91 91
FT CONFLICT 383 383
SQ SEQUENCE 437 AA; 49552 MW; 7E10218A9EE31360 CRC64;

alignment_scores:
Quality: 73.00 Length: 131
Ratio: 1.197 Gaps: 6
Percent Similarity: 46.565 Percent Identity: 23.664

alignment_block:
us-09-823-101-3 x VIPs_RAT ..
Align seg 1/1 to: VIPs_RAT from: 1 to: 437

57 TCGTTGGCTGGATCTTCTGTGAGCTCTGCTGATGCATGCTGCTAG 106
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
9 CysTyrCysTrpLeuValArgValSerSerIleHisProGluCysAr 25
107 CGTATCTTAATTTCTGCACAGTCATATGATGACAGACACAGTGGC 156
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
25 gphenHisLeuGluIleGlnGlu.....GluGluThrLysCysAa 38
157 TGGACAGTGTGGCATGCACTTTCACGATGACGTGGTCACCAACATGAG 206
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
38 IagIleuLeuSerSerGlnMetGlu.....Asn 47
207 CACACCTGTGTGTCACANTGTTGACAGAGTATACACACAGTGC 256
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
48 HisAlaGalaCys.....SerGlyValTyrPaspAsnIleTh 59
257 NTGTGNGAGCTATCCCGCATGGGGGAGACATGATTTTCCCAAAAT 306
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
59 rCysTrpArgProAlaAspIleGlyGluThrValThrValProCysProL 76
307 TTATCCACCCACCAATTTATTAACAAGAGATTTTATATGAGGGGGG 356
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
76 yValAlaPheSerAsn.....PheTyrSerArg..... 84

```

357 GGAGAGACAGCGACACAGGAATATACAGAG.....GAGATGG 397
 85ProGlyAsnIleSerIysAsnGlyThrSerAspG1 96
 398 ATAAACGAAACAGAAATACAAAAACACATCGCGGTGGTTAT 440
 96 yTTPserGluThrPheProAspPheIleAspAlaCysGlyTyr 110

seq_name: SwissProt_39:08D2_HUMAN

seq_documentation_block: PRT; 311 AA.
 ID 08D2_HUMAN STANDARD; PRT; 311 AA.
 AC Q9GZM6;
 DT 20-AUG-2001 (Rel. 40, Created)
 DT 20-AUG-2001 (Rel. 40, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE OLFACTORY RECEPTOR 8D2 (OLFACTORY RECEPTOR-LIKE PROTEIN JCG2).
 GN OR8D2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 OX NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Tongue;
 RA Gaubin J.-C., Chobert J.-M., Haerle T.;
 RT "Molecular cloning of olfactory receptor-like mRNAs expressed in human tongue."
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: PUTATIVE ODORANT RECEPTOR.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN THE TONGUE.
 CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: AF162668; AAG43386.1; -;
 DR EMBL: AF162669; AAG43387.1; -;
 DR InterPro: IPR000276: GPCR_Rhodopsn.
 DR Pfam: PF00001: 7tm_1.1
 DR PRINTS: PR00237; GPCR_RHODOPSN.
 DR PROSITE: PS00237; G_PROTEIN_RECPE_F1_1; FALSE_NEG.
 DR PROSITE: PS50262; G_PROTEIN_RECPE_F1_2; 1.
 KW G-protein coupled receptor; Transmembrane; Glycoprotein;
 KW Multigene family; Olfaction.
 FT DOMAIN 1 25 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 26 49 1 (POTENTIAL).
 FT DOMAIN 50 57 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 58 79 2 (POTENTIAL).
 FT DOMAIN 80 100 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 101 120 3 (POTENTIAL).
 FT DOMAIN 121 139 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 140 158 4 (POTENTIAL).
 FT DOMAIN 159 195 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 196 219 5 (POTENTIAL).
 FT DOMAIN 220 236 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 237 259 6 (POTENTIAL).
 FT DOMAIN 260 272 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 273 292 7 (POTENTIAL).
 FT DOMAIN 293 311 CYTOPLASMIC (POTENTIAL).
 FT DISULFID 97 189 BY SIMILARITY.
 FT CARBOHYD 5 5 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT SEQUENCE 311 AA: 34857 MW: ECE606FBBC962A82 CRC64:

alignment_scores:

Quality: 72.50 Length: 178
 Ratio: 0.863 Gaps: 11
 Percent Similarity: 47.191 Percent Identity: 24.719

alignment_block:

us-09-823-101-3/rev x 08D2_HUMAN ..

Align seq 1/1 to: 08D2_HUMAN from: 1 to: 311

696 GGGCTATACCTG.....AGATTATTTTGT 671
 33 GlyIleuYrValAlaThrValAlaGlyAsnLeuGlyMetIlePheLeu 49
 670 TCTTCATTATAGTCGG.....CCCCACGAAATATTTCTCGGCGTC 627
 49 eAlaLeuSerSerGlnLeuYrProValTyrTyrPheLeuSerHisL 66
 626 TCCCTCATAGTCGATTAATACATTAATATTCCTGCTGCTGCC 577
 66 euserPheIleasp.....LeucYrSerSerValIleThrPro 79
 576TCCCTCCCCCACCACCAAAATAGTT..... 550
 80 LysMetLeuValAsnPheValProGluGluAsnIleIleSerPheLeuG1 96
 549TTGTCTTAATTTTGTCTTCTACT.... 523
 96 uCysIleThrGlnLeuYrPhePheLeuIlePheValIleAlaGluTyr 113
 522CTCTCTTATTTGCCCCC 505
 113 YrIleuLeuThrAlaMetGluTyrAspArgTyrValAlaIleCysArgPro 129
 504 CTTTAAATAAATATCTCTCTTTTCATATTGTAATCTCTCTGTC 455
 130 Leu.LeuTyrAsnIleValMetSerHisArgValCysSerIleMeteta 146
 454 CTAAATAAATAAATA.....ACACACCGC 429
 146 IValValIleTyrSerLeuGlyPheLeuThrAlaThrValHisThrThrArg 162
 428 ATGTGTTTTTGTATTTCTGT.....TTTCGTTATCATCTCTCTGT 385
 163 MetSerValLeuSerPheCysArgSerHisThrValSerHisThrPheCys 179
 384 TGAATTCTCTGTCCTGCTCTCTCTCC...CCCCCCCATGTA...A 341
 179 sAspIleLeuProLeuLeuThrLeuSerCysSerThrHisIleAsnG 196
 340 AATATCTCTGTTATTTATGTTGGGTGAT 309
 196 IuIleLeuPheIleIleGlyGlyValAsn 206

seq_name: SwissProt_39:VE5_RHPV1

seq_documentation_block: PRT; 157 AA.
 ID VE5_RHPV1 STANDARD; PRT; 157 AA.
 AC P24834;
 DT 01-MAR-1992 (Rel. 21, Created)
 DT 01-MAR-1992 (Rel. 21, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE PROBABLE E5 PROTEIN.
 GN E5.
 OS Rhesus papillomavirus type 1 (Rhpv 1).
 OC Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;
 OC Papillomavirus.
 OX NCBI_Taxid=10570;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91135018; PubMed=1847267;
 RA Ostrow R.S., Labresh K.V., Fairs A.J.;
 RT "Characterization of the complete Rhpv 1 genomic sequence and an integration locus from a metastatic tumor."

```

RL   Virology 181:424-429(1991).
CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its use
CC by non-profit institutions as long as its content is in no way used for
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR   EMBL; M60184; AAA79315.1; ALT_SEQ.
DR   PIR; F38503; W5WLR1.
KM   Early protein.
SO   SEQUENCE 157 AA; 17398 MW; AC7AA67158844686 CRC64;

alignment_scores:
      Quality:    72.00          Length:    73
      Ratio:      2.000         Gaps:       4
Percent Similarity: 49.315     Percent Identity: 26.027

alignment_block:
US-09-823-101-3 x VE5_RHPV1 ..

Align seg 1/1 to: VE5_RHPV1 from: 1 to: 157

  9 GTGCTGCACGTGGTTCTGCTGT.....GGACCCTATGC 43
   :|||::: |||::: |||::: |||::: |||:::
 24 LeucylsAlaIasnCysCysrYrAlaclysProProbiohecy 40
   ::|||::: |||::: |||::: |||::: |||:::
 44 CTGTCAACAGGTGTGTCCTTGCGATCTTCTTGTAAGTCTG.... 87
   :|||::: |||::: |||::: |||::: |||:::
 40 sCyspheTrpleucysPheCyscyscSphE.CysLeuAlaLeucysPhe 56
   .....CTATGCATCCTGGT 101
   ::|||::: |||::: |||::: |||::: |||:::
 57 ValHisleuLeuSerArgCysPheCysValPheProvalCysLeuSerVa 73
   .....
102 CCTAAGCATTCCTATAATTCTCAACAGTCATATCGATGCAGACAGCAA 151
   | ||||| ::::::::::: ||| ::::::: |||
 73 lAlaIatyrAlaValalValleuglYalHis...SerGIuProvalCys 89
   152 GTGCTGCAGACTGTGG 168
   || |||::: |||:::
 89 erPheTrpserValphe 94

seq_name: Swissprot_39:ERD1_ARATH

seq_documentation_block:
ID ERD1_ARATH STANDARD; PTR; 945 AA.
AC P42762;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE ERD1 PROTEIN PRECURSOR.
GN ERD1.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxId=3702;
RX MEDLINE=94071876; PubMed=7504470;
RN SEQUENCE FROM N.A.
RP STRAIN=CV. COLUMBIA;
RA Kiyosue T., Yamaguchi-Shinozaki K., Shinozaki K.;
RT "Characterization of cDNA for a dehydration-inducible gene that
RT encodes a Clp A, B-like protein in Arabidopsis thaliana L.";
RL Biochem. Biophys. Res. Commun. 196:1214-1220(1993).
CC -!- FUNCTION: MAY INTERACT WITH A CLPP-LIKE PROTEASE INVOLVED IN
CC DEGRADATION OF DENATURED PROTEINS IN THE CHLOROPLAST.
CC -!- SUBCELLULAR LOCATION: CHLOROPLAST (POTENTIAL).

```

```

CC -1- INDUCTION: BY DEHYDRATION STRESS. INDUCED AFTER ONE HOUR OF
CC DEHYDRATION-STRESS AND REACHES MAXIMAL LEVELS AFTER 10 HOURS.
CC -1- SIMILARITY: BELONGS TO THE CLPA/CLPB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: D17582; BAA04506.1; -.
CC InterPro: IPR001939; AAA_subfam.
CC DR InterPro: IPR001270; CLP_AB.
CC DR Pfam: PF00004; AAA; 1.
CC DR PRINTS: PRO0300; CLP_POTFEASEA.
CC SMART: SMO0382; AAA; 2.
CC DR PROSITE: PS00870; CLPAB_1; 1.
CC DR PROSITE: PS00871; CLPAB_2; 1.
CC KW Chaperone; ATP-binding; Repeat; Chloroplast; Transit peptide.
CC FT TRANSIT 1 945 CHLOROPLAST (POTENTIAL).
CC FT CHAIN ? 945 ERD1 PROTEIN.
CC FT DOMAIN 271 523 I.
CC FT NP_BIND 316 781 ATP (POTENTIAL).
CC FT NP_BIND 664 671 ATP (POTENTIAL).
CC SQ SEQUENCE 945 AA; 103234 MW; 11EF233ZC78F656B CRC64;

Alignment_scores:
    Quality:      72.00      Length:      183
    Ratio:         0.867      Gaps:        11
    Percent Similarity: 45.355   Percent Identity: 23.497

alignment_block:
US-09-823-101-3 x ERD1_ARATH ..

Align seg 1/1 to: ERD1_ARATH from: 1 to: 945

133 ATCGATGACAGACAGCAAGCAAGTCGGTGGAACGTGGGCATCGAGTTCCAGCA 182
||||| ::: ::::::::::::::|||
132 TLeasPlYsAlaIarGluaIValItPrSerLTetTPaspGluaIAsmsE 148
183 TGACCTGGTCAACCAACATGATGAGCACACAATTGTGTGCCACANTGGTTT 232
:|||| :::::::::: |||:::||
148 rAspeerLYsgIngluaIsaSerthr 158
233 GAACAGAGTATAGCACACACGTCACNTGTGNGAGACTATCCCCGATGGGG 282
159 ..... "||||| :::: SerLyserlySerthr 164
283 GACATGAGTTTTCCCACCAAAATTATTCACCACCACCATTAATAACAA 332
||||| ||||| |
165 AsmPePrOPheSer..... lIeSerThrly 173
333 GAGATATTATTATGAGGGGGGGGAGAGACGAGGACACAGCAAAATAT 382
|||| | | :::::::::: || |
173 sArgYaLPHe..... GluaIaIaValIGluTYrs 183
383 CAACAGAAGAATGATATAAACGAAAAACAGAAATAC..... AAAAAACAC 426
|| | | ||||| ::::| |
183 eR...ArghrMetAsp..... cysGLnTyrlleAlaIaPrOgLuHIS 195
427 ATGCCCGTGTATTATTTTTTTA..... TTGGCAGC..... 459
::: ||| ::::| | |||||
196 tleaIaVGIyLeuPhenrThVaIspspGIySeRaIaGIaYrValle 212
460 .AGGAGATTACAATATGAAAAAGAAAGAGTAGTAATTATTT 498
::: ||||| ::::| |
212 uLysrArgLeu..... gLYaIaAsmMetasnLeuThrAlaIa 226

```

499TTAAGGGGGGGAATAGAGAGAGTACAGAAACAAA 537
 226 lalEurThrArgLeuYsGlySerPheGluLeuAlaLysAspGlyArgGluProSer 242
 538 TTAAGACAAACAACTATTTT.....TTTGG 563
 243 SerSerSerIysGlySerPheGluSerProSerGlyArgIleAlaG1 259
 564 TGGGGGGGAGAGAGAGACAGACAGAGATATATTATGATTATTC 612
 259 YSerGlyProGlyGlyLysAlaLysAlaValLeuGluInPheCys 275

seq_name: SwissProt_39:ABC3_HUMAN

seq_documentation_block:
 ID ABC3_HUMAN STANDARD; PRT; 1704 AA.
 AC 099758; 092473;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DE ATP-BINDING CASSETTE, SUB-FAMILY A, MEMBER 3 (ATP-BINDING CASSETTE
 DE TRANSPORTER 3) (ATP-BINDING CASSETTE 3) (ABC-C TRANSPORTER).
 GN ABC3 OR ABC3.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Thyroid carcinoma;
 RX MEDLINE=96326608; PubMed=8706931;
 RA Klingbauer N., Hofmann F.;
 RT "Primary structure of a novel ABC transporter with a chromosomal
 RT localisation on the band encoding the multidrug resistance-associated
 RT protein.";
 RL FEBS Lett. 391:61-65(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97179225; PubMed=9027511;
 RA Connors T.D., van Raay T.J., Petry L.R., Klinger K.W., Landes G.M.,
 RA Burn T.C.;
 RT "The cloning of a human ABC gene (ABC3) mapping to chromosome
 RT 16p13.3.";
 RL Genomics 39:231-234(1997).
 CC -1- FUNCTION: MAY BE A TRANSPORTER, ITS NATURAL SUBSTRATE HAS NOT BEEN
 CC FOUND YET (BY SIMILARITY). MAY ACT AS AN EFFLUX PUMP FOR
 CC CHEMOTHERAPEUTIC DRUGS.
 CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN LUNG, FOLLOWED BY BRAIN,
 CC PANCREAS, SKELETAL MUSCLE AND HEART. WEAKLY EXPRESSED IN PLACENTA,
 CC KIDNEY AND LIVER. ALSO EXPRESSED IN MEDULLARY THYROID CARCINOMA
 CC CELLS (MTC) AND IN C-CELL CARCINOMA.
 CC -1- DOMAIN: MULTIFUNCTIONAL POLYPEPTIDE WITH TWO HOMOLOGOUS HALVES,
 CC EACH CONTAINING AN HYDROPHOBIC MEMBRANE-ANCHORING DOMAIN AND AN
 CC ATP BINDING CASSETTE (ABC) DOMAIN (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY
 CC (ABC TRANSPORTERS). MDR SUBFAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL: U78735; AAC50967.1;
 CC EMBL: X97187; CAA65825.1;
 CC MIM: 601615;
 CC InterPro: IPR003593; AAA.
 CC InterPro: IPR003439; ABC_transportr.
 CC InterPro: IPR001687; ATP_GTP_A.
 CC Pfam: PF00005; ABC_tran; 2.
 CC SMART: SM00382; AAA; 1.

DR PROSITE; PS00211; ABC_TRANSPORTER; 1.
 KW ATP-binding; Transport; Transmembrane.
 FT TRANSMEM 22 42 POTENTIAL.
 FT TRANSMEM 249 269 POTENTIAL.
 FT TRANSMEM 307 327 POTENTIAL.
 FT TRANSMEM 344 364 POTENTIAL.
 FT TRANSMEM 373 393 POTENTIAL.
 FT TRANSMEM 405 425 POTENTIAL.
 FT TRANSMEM 447 467 POTENTIAL.
 FT TRANSMEM 925 945 POTENTIAL.
 FT TRANSMEM 1100 1120 POTENTIAL.
 FT TRANSMEM 1144 1164 POTENTIAL.
 FT TRANSMEM 1183 1203 POTENTIAL.
 FT TRANSMEM 1213 1233 POTENTIAL.
 FT TRANSMEM 1245 1265 POTENTIAL.
 FT TRANSMEM 1306 1326 POTENTIAL.
 FT NP_BIND 566 573 ATP (POTENTIAL).
 FT NP_BIND 1416 1423 ATP (POTENTIAL).
 FT CONFLICT 36 36 P -> S (IN REF. 2).
 FT CONFLICT 196 196 L -> P (IN REF. 2).
 SQ SEQUENCE 1704 AA; 191387 MW; AF0098DAF7A04F5F CRC64;

alignment_scores:
 Quality: 72.00 Length: 116
 Ratio: 1.263 Gaps: 7
 Percent Similarity: 49.138 Percent Identity: 28.448

alignment_block:

US-09-823-101-3/rev x ABC3_HUMAN ..

Align seq 1/1 to: ABC3_HUMAN from: 1 to: 1704

690 TACTTGAGATTATTTTGTCTTCATTATAGTCGGCCCCCAATAATA 641
 309 PheLeuLeuPhePheLeuPheLeu..... 316
 640 TTCTCTCGGCGTCCCTCATAGCTGCATATACTATAAT.....T 597
 317LeuIleAlaIleSerPheMetThrLeuLeuP 327
 596 ATTCTCTCGTGTCTCTCCCTCC.....TCCCCCCC 565
 327 hecysValIysValIysProAsnValAlaValLeuSerArgAspPro 343
 564 ACCAAAAATAGTGTGTCTTATTTTGTGTCTTCTACCTCTCTT 515
 344 SerLeuValLeuAlaPheLeuLeuCysPheAlaIleSerThrIleSer.. 359
 514 ATTCCCCCCCCCTTAATAAATAATTACTCTCTTTCATATTTGTAAT 465
 360 .PheSerPheMet.....ValSerThrPhePheSer..... 369
 464 CTCCTCTGCTGCTAATAAAAAAATAACACAGCGCATGTGTTTGTGA 415
 370LysAlaAsnMetAlaIleAlaIlePheGly 379
 414 TTTCGTGTTTCGTTATCATCTTCTTGTGATTTCTCTGTCCTCC 367
 380 PheLeuYrPhePheThrTyrlleProTyr...PhePheValAlapro 394

seq_name: SwissProt_39:HYPA_HYPLI

seq_documentation_block:
 ID HYPA_HYPLI STANDARD; PRT; 256 AA.
 AC P35587; Q25080;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE HYPODERMIN A PRECURSOR (EC 3.4.21.-) (HA).
 OS Hypoderma lineatum (Early cattle grub) (Common cattle grub).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Oestridea; Oestrideae; Hypoderma.
 OX NCBI_TaxID=7389;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95107346; PubMed=7808473;
 RA Moire N., Bigot Y., Perleuet G., Boulard C.;
 RT "Sequencing and gene expression of hypodermis A, B, C in larval
 stages of Hypoderma lineatum."
 RL Mol. Biochem. Parasitol. 66:233-240(1994).
 RN [2]
 RP SEQUENCE OF 3-256 FROM N.A.
 RA Kuhn I., Files J.G., Pruett J.H., Temeyer K.B.;
 RL Submitted (XXX-1993) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 31-51.
 RX MEDLINE=81232922; PubMed=7018579;
 RA Tong N.T., Imhoff J.M., Lecroisey A., Kell B.;
 RT "Hypodermin A, a trypsin-like neutral proteinase from the insect
 Hypoderma lineatum."
 RL Biochim. Biophys. Acta 658:209-219(1981).
 CC -1- FUNCTION: SPECIFICITY, LIMITED TO CARBOXYL SIDE OF ARGININE
 CC RESIDUE IN B-CHAIN OF INSULIN.
 CC -1- SUBCELLULAR LOCATION: SECRETED.
 CC -1- DEVELOPMENTAL STAGE: LARVAL-SPECIFIC.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 CC TRYPSIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: X74303; CAAS2356.1; -;
 DR EMBL: L24814; AAA29224.1; -;
 DR PIR: A21590; A21590.
 DR HSSP: P00763; IDPO.
 DR MEROPS: S01.111; -;
 DR InterPro: IPR001314; Chymotrypsin.
 DR InterPro: IPR001254; Trypsin.
 DR InterPro: IPR00089; Trypsin.
 DR Pfam: PF00089; Trypsin; 1.
 DR PRINTS: PR00722; CHYMOTRYPSIN.
 DR SMART: SM00200; TRYP-SPC; 1.
 DR PROSITE: PS50240; TRYPSIN_DOM; 1.
 DR PROSITE: PS00134; TRYPSIN_HIS; 1.
 DR PROSITE: PS00135; TRYPSIN_SER; 1.
 KW Hydrolyase; Serine protease; Zymogen; Signal.
 FT SIGNAL 1 22
 FT PROPEP 23 30
 FT CHAIN 31 256
 FT ACT_SITE 71 71
 FT ACT_SITE 116 116
 FT ACT_SITE 210 210
 FT DISULFID 56 72
 FT DISULFID 180 197
 FT DISULFID 206 230
 FT SITE 204 204
 FT SITE 37 37
 FT CONFLICT 45 45
 FT CONFLICT 51 51
 FT CONFLICT 248 248
 SO SEQUENCE 256 AA; 27907 MW; 63D12D337F66AA61 CRC64;
 alignment_scores:
 Quality: 71.50 Length: 87
 Ratio: 1.833 Gaps: 4
 Percent Similarity: 44.828 Percent Identity: 24.138
 alignment_block:
 US-09-823-101-3/rev x HYPA_HYPLI ..

Align seg 1/1 to: HYPA_HYPLI from: 1 to: 256
 354 CCCCCCATATATAATATCTCTTTATTATATGTTGGGTGATTAAT 305
 |||||:::|||||:::|||||:::|||||
 142 PROPRGIIuTYrAlaSpAlaIleValSerGlyTP..... 153
 304 TTTTGGGGAANAACATCATGCCCCCATCGGAGATGCTCN..... 262
 |||||:::|||||:::|||||:::|||||
 154GlyGIuThrLeuValProProGlyIleProAspGlnLeuA 168
 261CACANGTCACGTGTGCT..... 244
 168 rGSeValAspVallylIleIleHisArgIuIuScysAlaSerArgAsn 184
 243ATACCTGTTCAACACANG 224
 185 pHeGlyTYrGlySerAsnIleLysAlaSerMetIleCysAlaIleAlaI 201
 223 TGGACACACACACTGT.....GCTCATGTTGGTTGACC 189
 |||||:::|||||:::|||||:::|||||
 201 eGlyLysAspSerCysGlnGlyAspSerGlyProLeuValAlaAsnA 218
 188 ACCTCATCTGTG 178
 :::::|||||
 218 snLeuVal 221
 seq_name: SwissProt_39:YKR9_YEAST
 seq_documentation_block:
 ID YKR9_YEAST STANDARD; PRT; 679 AA.
 AC P34237; P34236;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-JUN-1994 (Rel. 19, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE HYPOTHETICAL 77.5 KDA PROTEIN IN RPL17A-STE3 INTERGENIC REGION.
 GN YKL179C.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 CC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94205264; PubMed=8154185;
 RA Wiemann S., Voss H., Schwager C., Rupp T., Stegemann J.,
 RA Zimmermann J., Grothues D., Sensen C., Erfle H., Hewitt N.,
 RA Banerji A., Ansoorge W.;
 RT "Sequencing and analysis of 51.6 kilobases on the left arm of
 RT chromosome XI from Saccharomyces cerevisiae reveals 23 open reading
 RT frames including the FAS1 gene."
 RL Yeast 9:1343-1348(1993).
 RN [2]
 RP REVISIONS:
 RA Wiemann S., Voss H., Schwager C., Rupp T., Stegemann J.,
 RA Zimmermann J., Grothues D., Sensen C., Erfle H., Hewitt N.,
 RA Banerji A., Ansoorge W.;
 RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: SOME TO HUMAN CENP-E, AND TO YEAST USO1 AND S. POMBE
 CC SPOC364.04C.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: X74151; CAAS2259.1; -;
 DR EMBL: Z28179; CAAB2021.1; -;
 DR PIR: S38011; S38011.
 DR SGD: S0001662; YKL179C.
 KW Hypothetical protein.

THIS PAGE BLANK (USPTO)

OM of: US-09-823-101-3 to: SPTREMBL_17:* out_format : pfs
Date: Jan 17, 2002 4:07 PM

About: Results were produced by the Gencore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL=framer.n2p.model -DEV=xlp
-Q=/cqr2.1/USPTO.spool/US09823101/runat_17012002.155138.22043/app-query.fasta_1.2239
-DB=SPTREMBL_17 -QFMT=fastan -SUFFIX=n2p.rspt -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPEL=0.000 -LOOPEXT=0.000
-OCAPOP=4.500 -OCAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FCAPOP=6.000 -FCAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blomsun62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=Pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=EXT -MINLEN=0 -MAXLEN=200000000
-USPR=US09823101.ecgn1_1_384 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPPY
-WAIT -THRAIDS=1

Search information block:

Query: US-09-823-101-3
Query length: 698
Database: SPTREMBL_17:*
Database sequences: 473505
Database length: 146272329
Search time (sec): 224.230000

```
score_list:
Sequence      Strd Orig      ZScore      EScore len  Documentation
SP_Invertebrate:018560 - 90.50 160.08 0.2185 555 018560 vargula hilgendorffii (
SP_Invertebrate:026195 - 88.50 163.25 0.4041 200 026195 plasmodium vivax, pval
SP_plant:Q9M1P9 - 86.00 153.93 0.7162 373 09M1P9 arabidopsis thaliana (mc
SP_Invertebrate:09XW59 - 85.00 151.52 0.9183 366 09XW59 caenorhabditis elegans
SP_Invertebrate:Q9N8H0 - 82.50 136.22 1.49 1737 09N8H0 trypanosoma brucei, f
SP_plant:Q9AW73 - 82.00 144.99 1.96 430 09AW73 guillardia theta (cryptic
SP_plant:Q9LR94 - 82.00 142.07 1.87 753 09LR94 arabidopsis thaliana (mc
SP_Bacteria:051099 - 82.00 140.89 1.84 773 051099 borrelia burgdorferi (ly
SP_plant:Q9Z5T9 - 81.00 140.09 2.42 652 09Z5T9 arabidopsis thaliana (mc
SP_virus:Q91BR8 - 80.00 140.51 3.23 462 091BR8 spodioplera exigua nucleoc
SP_Bacteria:0950U3 - 77.50 136.11 6.16 426 0950U3 shigella sonnei. orf15.
SP_Bacteria:Q9F740 - 77.50 136.11 6.16 426 09F740 shigella sonnei. wzx. 6/
SP_Invertebrate:Q9NG78 - 77.50 131.79 5.78 791 09NG78 caenorhabditis elegans
SP_Invertebrate:062186 - 77.00 135.85 7.07 384 062186 caenorhabditis elegans
SP_fungi:Q9P472 - 77.00 134.40 6.92 472 09P472 emeticella nidulans (ast
SP_plant:Q9PFT7 - 76.50 134.45 7.98 407 09PFT7 arabidopsis thaliana (mc
SP_Bacteria:035044 - 76.50 134.13 7.94 426 035044 shigella sonnei. form 1
SP_virus:Q9DMW2 - 76.00 137.72 9.66 221 09DMW2 rat cytomegalovirus (str
SP_plant:Q9FRV9 - 76.00 136.92 9.54 248 09FRV9 calystegia soldanella. f
SP_Bacteria:09S521 - 76.00 134.51 9.21 350 09S521 escherichia coli. putati
SP_rodent:Q9RIY7 - 76.00 128.48 8.42 829 09RIY7 mus musculus (mouse). ad
SP_plant:Q9SC36 - 75.50 137.22 11.05 206 09SC36 pisum sativum (garden pe
SP_Invertebrate:Q9VD26 - 75.50 134.62 10.63 299 09VD26 drosophila melanogaste
SP_Bacteria:Q9S0Y3 - 75.00 131.22 10.26 421 09S0Y3 plesiomonas aeruginosa.
SP_Bacteria:Q9HWT2 - 75.00 131.27 11.66 419 09HWT2 pseudomonas aeruginosa.
SP_fungi:Q9HGU6 - 75.00 130.42 11.52 471 09HGU6 blumeria graminis. prote
SP_fungi:Q9H077 - 75.00 130.33 11.50 480 09H077 aspergillus niger. camp
SP_Bacteria:084150 - 75.00 129.93 11.43 507 084150 shigella flexner. camp
SP_virus:Q9WS53 - 75.00 123.61 10.73 933 09WS53 oryza sativa (rice). hyf
SP_human:Q14674 - 75.00 121.09 10.44 1217 09WS53 simian t-cell lymphotr
SP_fungi:Q9U5F9 - 74.50 129.44 13.03 472 09U5F9 erysipe graminis (subst
SP_Bacteria:Q9PKP0 - 74.50 128.95 12.99 506 09PKP0 chlamydia muridarum. mor
SP_plant:Q9STN1 - 74.00 130.54 15.33 350 09STN1 arabidopsis thaliana (mc
SP_Bacteria:004532 - 74.00 125.12 14.14 760 004532 arabidopsis thaliana (mc
SP_Bacteria:Q913K3 - 74.00 124.83 14.08 792 0913K3 rummocooccus flavifaciet
SP_rodent:Q9ERP5 - 73.50 130.30 17.60 314 09ERP5 mus musculus (mouse). b2
SP_Invertebrate:Q9N474 - 73.50 129.96 17.51 330 09N474 caenorhabditis elegans
SP_Invertebrate:Q9GR44 - 73.50 129.37 17.36 359 09GR44 toxoplasma gondii. mid
SP_Invertebrate:044330 - 73.50 128.82 17.22 388 044330 manduca sexta (tobacco
```

```
SP_Bacteria:Q31347 - 73.50 128.52 17.15 405 031347 bacillus cereus. orf1
SP_human:Q9UJY2 - 73.50 127.69 16.94 456 09UJY2 homo sapiens (human).
SP_human:Q9H6L0 - 73.50 126.50 16.64 541 09H6L0 homo sapiens (human).
SP_plant:Q9SKD1 - 73.50 125.10 16.30 661 09SKD1 arabidopsis thaliana
SP_human:Q9Y4G1 + 73.50 122.23 15.62 997 09Y4G1 homo sapiens (human).
```

seq_name: sp_invertebrate:018560

seq_documentation_block:

```
ID 018560 PRELIMINARY; PRT; 555 AA.
AC 018560;
DT 01-JAN-1998 (TRMBREL. 05 Created)
DT 01-JAN-1998 (TRMBREL. 05, last sequence update)
DT 01-JUN-2001 (TRMBREL. 17, last annotation update)
DE LOCIFERASE.
OC Vargula hilgendorffii (Sea firefly) (Cypridina hilgendorffii).
OC Eukaryota; Metazoa; Arthropoda; Crustacea; Ostracoda; Myodocopa;
OC Myodocopa; Cypridinoida; Cypridinidae; Vargula.
CX NCBL_TaxID=6674;
[1]
RN
RP SEQUENCE FROM N.A.
RX MEDLINE=97250550; PubMed=9096406;
RA Miesenböck G., Rothman J.E.;
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; 089490; AAB86460.1; -.
DR InterPro; IPR001846; Vwd.
DR Pfam; PF00094; vwd; 2.
DR SMART; SM00216; VWD; 2.
SQ SEQUENCE 555 AA; 61583 MW; CE12E56060877863 CRC64;
```

alignment_scores:

```
Quality: 90.50 Length: 134
Ratio: 1.775 Gaps: 6
Percent Similarity: 38.060 Percent Identity: 28.358
```

alignment_block:
US-09-823-101-3/rev x 018560 ..

Align seg 1/1 to: 018560 from: 1 to: 555

```
372 TGTCCGTCCTCTCCCTCCCTCATATAATATCTTGTATTAA 323
|||||
22 CysProvaIgluAgluProProser..... 30
322 TTGTGTGGGTGATTAATTTTGGGAAACATCATGCCCCCATCG 273
|||||
31 .....SerthProthrValProthrSerC 39
272 GGATGACTCCNAC.....ANGTCAGTTGTGCTATA..... 241
|||
39 ysgluAlaLysgluLysgluLysgluLysgluLysgluLysglu 55
240 .....CTCTGTTCAACACANTGTGGACACACAG 212
56 ArgaspIleLeuSerAspLysLysLysgluLysgluLysgluLysglu 72
211 TTGTGTGTCATGTTGTTGACACAGTCATGCTGAACATGACACAG 162
|||
72 sCys..... 73
161 TTCACGACATGCTGTCATGATGATGATGATGATGATGATGATGAT 112
|||
74 .....ArgmetCysgluLysgluLysgluLysgluLysgluLysglu 88
111 ATACGCTAGACACAGATGATGATGATGATGATGATGATGATGATG 62
|||||
|||||
```

88 yTyrPheArgTInrPheTyGlyLysArgPheAsnDheGlnGluProGlyL 105
 61 AAGCACCTGGTGCACAGCATAGGTCACAGCAACCAACCATGGACGA 12
 105 ystyValLeuAlaArgGly.....ThrlsGlyGlyAspIrrpSer 118
 seq_name: sp_invertebrate:Q26195

seq_documentation_block:
 ID Q26195 PRELIMINARY; PRT; 200 AA.
 AC Q26195;
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)
 DE PVAL GENE.
 GN PVAL.
 OS Plasmodium vivax.
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=5853;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=INDIAN;
 RX MEDLINE=98171310; PubMed=9512356;
 RA Dhar A., Gupta S., Sharma Y.D.;
 RT "Alu elements in a Plasmodium vivax antigen gene."
 RL FEMS Lett. 423:193-197(1998).
 DR EMBL; X92485; CAA63219.1; -
 SO SEQUENCE 200 AA; 22833 MW; 40EF6A799FELC73 CRC64;

alignment_scores:
 Quality: 88.50 Length: 212
 Ratio: 0.932 Gaps: 12
 Percent Similarity: 44.811 Percent Identity: 26.415

alignment_block:

US-09-823-101-3/rev x Q26195 ..

Align seg 1/1 to: Q26195 from: 1 to: 200

```

687 TTGAGATTATTTGTTCTCATTTATGCGCCGCCACTAAATATT 638
||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:
3 LeuserPhePheLeuPheLeuPheLeuSerPheLeuProIrrPheLeuPh 19
637 CTCGCGGCTCTCCCTCATAGCTGCATATACTAATAATATCTCG 588
|||:|||||:|||||:|||||:|||||:|||||:|||||:
19 eSerSerPheLeuProSerPheProSer.....PheHisArgSerMetI 34
587 TCGGTCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCT 538
: : : : : : : : : : : : : : : : : : : : : : : : :
34 LeGlnLysProSer.....Val.ArgLeuHisArg 43
537 TTTTGTGTTTCTACTCTCTCTATTCCTCCCTCCCTCCCTCCCT 488
: : : : : : : : : : : : : : : : : : : : : : : : :
43 GLeuCySPheProSerLeuPro...SerProLeuLeuSerProLeuL 59
487 TCTCT...TCTTTCAATATTTGTAATCTCTCTGCTAATAAAAAA 441
|||:|||||:|||||:|||||:|||||:|||||:|||||:
59 euAlaLeuSerPheProSerProSerProLeuProPheProSerVal 75
440 ATACCAACGCGCATGTGTTTGTATTTCTGTTTGTATTCATCATT 391
: : : : : : : : : : : : : : : : : : : : : : : : :
76 SerLeuSerLeuPheCySPheProSerPheSerPheLeu...Pro...S 91
390 CCTGTGTGATATTTCCTGTGCGCTCTCTCTCCCTCCCTCCCTCC 350
|||||:|||||:|||||:|||||:|||||:|||||:
91 erLeuPProPhePheProLeuLeuSerLeuProPheProLeuArgPhe 107
350 ..... 350
108 PheIlePhePhePheArgAspArgValLeuLeuCySHisProGlyTTPse 124
349 .....CTCATATATAA 340

```

124 rAlaValAlaInserLeuPheThrValAlaSerThrPheLeuValLysG 141
 339 ATATCTCTGTTATTAATGTTGGGCGTGATTAATTTTGGGCAAAAC 290
 :
 141 InSerSerCys.....LeuGlyLeuProSerSerTrpAspTyrArg 154
 289 TCATGTCCCTCCCATCGGGGATAGCTCCACACATGCTGTGCTATAC 240
 :
 155 ArgIlePro.....ProHisLeuAlaHisPheSerPhe 166
 239 TCTGTTCAACACANTGTGGACACAAAGTTGTGCTCATGTTGTTGAC 190
 :
 166 heCySArgAsnLys.....SerLeuThrMetLeuPro 176
 189 CACGTCAATCGTGAACACTGCATCCACACGTT 160
 :
 177 ArgLeuIleLeuAsnSerTrpProGlnVal 186
 seq_name: sp_plant:Q9MIF9

seq_documentation_block:
 ID Q9MIF9 PRELIMINARY; PRT; 373 AA.
 AC Q9MIF9;
 DT 01-OCT-2000 (TREMBlrel. 15, Created)
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE HYPOTHETICAL 42.7 KDA PROTEIN.
 GN FRK21.60.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Jordan N., Bangert S., Wiedelmann R., Voss H., Unselid M., Mewes H.W.,
 RA Lemcke K., Mayer K.F.X., Quettier F., Salanoubat M.;
 RL Submitted (FEB-2000) to the EMBL/Genbank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RL Submitted (FEB-2000) to the EMBL/Genbank/DBJ databases.
 CC -1- SIMILARITY: CONTAINS A RING-TYPE ZINC FINGER.
 DR EMBL; AL138657; CAB75477.1; -
 DR InterPro: IPR002106; AA_trna_ligase_II.
 DR InterPro: IPR002867; IIR.
 DR InterPro: IPR001841; Znf_ring.
 DR Pfam; PF01485; IIR; 1.
 DR SMART; SM00184; RING; 2.
 DR PROSITE; PS00179; AA_TRNA_LIGASE_II_1; UNKNOWN_1.
 DR PROSITE; PS00518; ZINC_FINGER_C3HC4; 1.
 KW Hypothetical protein; zinc-finger.
 SQ SEQUENCE 373 AA; 42688 MW; A44845BF00667BB CRC64;

alignment_scores:
 Quality: 86.00 Length: 209
 Ratio: 0.977 Gaps: 10
 Percent Similarity: 42.105 Percent Identity: 22.967

alignment_block:

US-09-823-101-3/rev x Q9MIF9 ..

Align seg 1/1 to: Q9MIF9 from: 1 to: 373

```

648 ACTAATATATTTCTCGCGCTCTCCCTCATAGCTGCATATACTATA 599
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
70 ThrLysAlaValSerPheArgIleAsnHisIleSerMetPheAspSps 86
598 T.....TATTTCTGTGCTGTGCTCTCCCTCCCTCCCTCCCTCC 558
- : : : : : : : : : : : : : : : : : : : : : : : : :
86 nProGluIlePheGluLeuValMetGlyArgSerValProLysAspLysL 103

```


452ATAAAAAAAAAATACACGGCAGTGTGTTTGTAT 414
 225 leasnlyseuasnarglulysilleuPhepheketilleuLeuLyr 241
 413 TTCTGTTTCGTTATCCATCTTCCTGTGATATTCCGTGTCGCCGTC 364
 242leuasnleuPhe..... 245
 363 TCTCTCCCGCCCGCCATATATAATATCTGTTATTATTGTTGGG 314
 246SerProPheSerThrArgIlelleuLeuLysPhe.....L 259
 313 TGGATTAATTTTGGGGAATACTATGTCGCCCGCATCGGGATAGTTC 264
 259 euasnlyspheilleysasnAnPheLys.....IleAlaIlelle 273
 263 CACACAGTCAGCTGTGCTATCTGTTCAACCAANTGTGGACACACA 214
 274 LeuIleArgHisTyrIlePheTyrSerLysGlnThrAsnIleasnGlnAs 290
 213 AGTGTG 207
 290 nilelle 292

seq_name: sp_plant:09LR94

seq_documentation_block:

ID 09LR94 PRELIMINARY; PRT: 653 AA.

DT 01-OCT-2000 (TREMBlrel. 15, Created)

DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)

DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)

DE T23E23.16

OS Arabidopsis thaliana (mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.

OX NCBI_TaxID=3702;

RN [1]

RP SEQUENCE FROM N.A.

RA Ecker J.R.;

RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RA Cheuk R., Shinn P., Brooks S., Buehler E., Chao O., Johnson-Hopson C.,

RA Khan S., Kim C., Alcafi H., Bei B., Chin C., Chio J., Choi E.,

RA Conn L., Conway A., Gonzalez A., Hansen N., Howing B., Koo T., Lam B.,

RA Lee J., Lenz C., Li J., Liu A., Liu J., Liu S., Mukharisky N.,

RA Nguyen M., Palm C., Pham P., Sakano H., Schwartz J., Southwick A.,

RA Thaveri A., Toriumi M., Vaysberg M., Yu G., Davis R., Federspiel N.,

RA Theologis A., Ecker J.;

RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.

CC -! SIMILARITY: TO RNA-DIRECTED DNA POLYMERASE (REVERSE

CC TRANSCRIPTASE)

DR EMBL: AC002423; AAF87143.1; -

DR InterPro: IPR000477; RVTse.

DR Pfam: PF00078; Ivtc. 1.

KW RNA-directed DNA polymerase.

SEQUENCE 653 AA; 75802 MW; D7AEA9219C604814 CRC64;

alignment_scores:

Quality: 82.00 Length: 180
 Ratio: 0.965 Gaps: 11
 Percent Similarity: 47.222 Percent Identity: 23.889

alignment_block:

US-09-823-101-3 x 09LR94

Align seg 1/1 to: 09LR94 from: 1 to: 653

11 GTCCTCAGTGGG.....TTCCGTGCTGGACCCCTATGCTGTCACAGG 54
 511 leuSerThrGlyAspLysMetLeuGlnTrpAsnArgArgLeuSerProth 527
 55 T...GTGCTTTCCTGGAT.....CTTCTGTGGA 80
 527 rcysValleuCySAsnAsnAsnIleGlnThrArgAsnHisleuPhePheS 544
 81 GTCCTGCTATGCAATCCCTGCTCCTAGCGTATCTTAATTCACACAGTC 130
 544 ecyscysTyrThrAla..... 549
 131 ATATGATGACAGACAGACAGAGCGGTGGAGCTGTGCATGACATTCAC 180
 550 ...GlnIleTrpGlnLeuAlaIalysAsnIleTyrLysAlaIalysPheS 565
 181 GATGACGTGTCACACACATGAGACACACTGTGTGTCACACANTG. 229
 565 rhrAsnTrpSerThrIleLeuThrSer.....ValSerThrThrTrpA 580
 230TTGACACAGATATAGACACAGACGACNTGTGNGAGACTATC 271
 580 rGAsnArgThrGlnSerPheLeuAlaArgTyr..... 590
 272 CCCGATGGGGGACATGAGTTTTCGCCCAAAATTATCCACCOCACCA 321
 591IlePheGlnAlaThrIleHis..... 597
 322 ATTAATAACAGATATTTTATGAGGGGGGGGAGACAGACGCGAC 371
 598ThrIleTrpHisGlnArgAsnGlyArgArgHisGlyG 610
 372 ACAGAAATATGACACAGAGA.....TGATTAACGAAGAAC... 409
 610 IuArgSerAsnSerAlaThrHisLeuIleTrpTrpLeuAspLysGlnMet 626
 410 AGAAATACAAAAAACACATCGCGTGTATTTTATTTATTAAGCAGG 459
 627 ArgAsnGlnIleSerThrIleAla.....AlaSe 636
 460 AGCAGATTACAATATGAAGAAGAGAGACTAATTTATTTT 499
 636 rGlyAspHisArgTyrAspLysProLeuIleuTrpPhe 649

seq_name: sp_bacteria:051099

seq_documentation_block:

ID 051099 PRELIMINARY; PRT: 773 AA.

AC 051099;

DT 01-JUN-1998 (TREMBlrel. 06, Created)

DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)

DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)

DE HYPOTHETICAL 89.8 KDA PROTEIN.

GN BB0072.

OS Borrelia burgdorferi (Lyme disease spirochete).

OC Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.

OX NCBI_TaxID=139;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=ATCC 35210 / B31;

RX MEDLINE=98065943; PubMed=9403685;

RA Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A.,

RA Lathigra R., White O., Ketchum K.A., Dodson R., Hickey E.K., Gwinn M.,

RA Dougherty B., Tomb J.-F., Fleischmann R.D., Richardson D.,


```

229 eu.Cys...SerAlaCysGlyLeuAenGlyAspArgSerProTyrIleCy 244
251 GTTG...TGCATACCTCTGTTCAACCAACANTGTCGACACCAAGTTGCT 205
244 sValGlnCysAspPheMetIle.....HisGlnValCysL 256
204 CTCATGTTGTTGACCCATCATCGTACATGCATGCCACAC..... 163
256 euhIstIeu.....ProArg.ValIleAenIleAenArgHisAspHisar 270
162 .GTTCCAGCAGCACT.....T 150
270 gValSerArgThrSerValIeuGlyValValAsnSerValCysGlyValC 287
149 GCCTGTCTGTGATCATGATGACT 127
287 ysarGlnGlnysValAspTrpThr 294

```

seq_name: sp_virus:Q91BR4

seq_documentation_block:

```

ID Q91BR4 PRELIMINARY; PRT; 462 AA.
AC Q91BR4;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE ORF2 ORF1629.
OS Spodoptera exigua nucleopolyhedrovirus.
OC Viruses; dsDNA viruses, no RNA stage; Baculoviridae;
OC Nucleopolyhedrovirus.
OX NCBI_TaxID=10454;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93057353; PubMed=1431809;
RA Jukel W.F., van Strien E.A., Zuidema D., Goldbach R.W.,
  "Nucleotide sequence and transcriptional analysis of the polyhedrin
  gene of Spodoptera exigua nuclear polyhedrosis virus.";
RT J. Gen. Virol. 73:2813-2821(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=20036646; PubMed=10567663;
RA Jukel W.F., van Strien E.A., Heldens J.G., Broer R., Zuidema D.,
  Goldbach R.W., Vlak J.M.;
RT "Sequence and organization of the spodoptera exigua multicapsid
  nucleopolyhedrovirus genome.";
RL J. Gen. Virol. 80:3289-3304(1999).
RN [3]
RP SEQUENCE FROM N.A.
RA Zuidema D.;
RL Submitted (JUL-1992) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Jukel W.F.J., van Strien E.A., Heldens J.G.M., Broer R., Zuidema D.,
  Goldbach R.W., Vlak J.M.;
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF169823; AAF3533.1; -.
DR InterPro: IPR002965; P_rich_extensn.
DR PRINTS: PR01217; PRICHEXTENS.
SQ SEQUENCE 462 AA; 50358 MW; 21CF651DAF81CD CRC64;

```

alignment_scores:

Quality: 80.00 Length: 103
Ratio: 1.702 Gaps: 4
Percent Similarity: 45.631 Percent Identity: 26.214

alignment_block:

US-09-823-101-3/rev x Q91BR4 ..

Align seg 1/1 to: Q91BR4 from: 1 to: 462

653 CCCCACATAATATTCTCTCGGCGTCCCTC...TCATAGCTGCATAA 607

```

226 ProPromeSerThrSerIleValAlaSerProAspAspLysIleGlns 242
606 TACTAATATATTCTCTCGTCCGTCCTCCCTCCGCCACCAACAAAA 557
242 nysIleAspValValGlnThrAlaLeuProProProProPro.... 257
556 AATAGTTTGTCTTAATTATTTTGTTCCTACTCTCTCTTAATGCCCC 507
258 .....ProProProPro 261
506 CCGTTTAAAAATAAATTAATCTCTCTCTTTCATATTGTAATCTCTCT 457
262 PromeProSerSerIleIleSer.....IleProThrIleThrTh 275
456 GCCTAATATAAAAAAATAACACGCGCATGTGTTTGTATTCTGT 407
275 rPro..... 276
406 TTCGTTTATTCATCTTCCTGTGATATTTCCTGTGTCCTCTCTCTCC 357
277 .....ValAspGlnSerThrThrAsnAlaPheIle 286
356 CCCCCCGCT 348
287 ProProPro 289

```

seq_name: sp_bacteria:Q950U3

seq_documentation_block:

```

ID Q950U3 PRELIMINARY; PRT; 426 AA.
AC Q950U3;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE ORF45.
OS Shigella sonnei.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Shigella.
OX NCBI_TaxID=624;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HM383;
RA Chida T., Okamura N., Yoshida Y., Ohtani K., Arakawa E., Watanabe H.;
RT "Complete DNA sequence of the O-antigen (rfb) gene cluster in
  Plesiomonas shigelloides serotype 017 having the same O-antigen as
  Shigella sonnei: comparison with that of S. sonnei.";
RL Submitted (May-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=HM383;
RX MEDLINE=99036814; PubMed=9817819;
RA Houny H.H., Venkatesan M.M.;
RT "Genetic analysis of Shigella sonnei form I antigen: identification of
  novel IS630 as an essential element for the form I expression.";
RL Microb. Pathog. 25:165-173(1998).
DR EMBL: AB028135; BAA85073.1; -.
DR InterPro: IPR002797; Polysacc_synth.
DR Pfam: PF01943; Polysacc_synth; 1.
SQ SEQUENCE 426 AA; 49307 MW; 30072A02689F7195 CRC64;

```

alignment_scores:

Quality: 77.50 Length: 190
Ratio: 0.891 Gaps: 7
Percent Similarity: 45.789 Percent Identity: 22.632

alignment_block:

US-09-823-101-3/rev x Q950U3 ..

Align seg 1/1 to: Q950U3 from: 1 to: 426

677 TTTTGTCTTCATTTAGTCCGCCCCACCACTAATATTCTCTCGGCGT 628

```

93 PheMetPheTySerTyValSerAspSerIleSerLeuThrLeuVal 109
627 CTCCCTCATAGCTGCATATACTATATATCTCTCGCGTCCTCC 578
109 lTyrlIeSer.SerAlaTyRpheValCysIleSerSerGlyArgPheSer 125
577 CTCTCCCGCCCGCCACCAAAA.....AATAGTTT 549
126 LeuLeuGlnAlaValAlGlyArgPheArgCysGluLeuTyrlIeAsnIleTy 142
548 TTGTTCTTAATTTTGT.....TTCTACTCTCTCTTATTC 511
142 rSerThrIleTyrlIeGlyCysAsnLeuPheLeuSerLeuPheIleG 159
510 CCCCCCTTTAAATAATTAATCTCTCTTTTCATATTTGTATCTCC 461
159 lupro.....LeuTyrlTySer..... 164
460 TCCTGCCTAATAAATAAATAAACAACGCGCATGTGTTTGTATTTTC 411
165 .....AlaIleSe 167
410 TGTGTTGCTTATCCATCTCTCTGTGATATTCCTGTCTCTCT 361
167 rIlePheIleTySerIleSerIleLeuValPheSerSerHisLysCysA 184
360 CTCCCGCCCGCCCT.CATATAAA..... 339
184 snValProCysPheHisIleLysArgProSerIleLeuValTyrlLysAsp 200
338 .....TATCTCTGTATTAAT 322
201 PheLeuAspAlaThrProPheAlaIleLeuValLeuLeuAsnValIle 217
321 TGTGCGGCTGATATAATTTTGGGGGAAAAACTCATGTCCCGCATCGGG 272
217 uSerSerIleAspLeuPheIleLeuLysGluTyRPheserTyrlAsn... 232
271 GATAGCTCCNACANGTCACGTGCTATCTGTCAACCAACANTGTG 222
233 .....SerValAlaIleTyRglnValIleThrArgVal 243
221 GACACACAAGTTGTGCTC 204
244 AsnThrGlyLeuIleIle 249

```

seq_name: sp_bacteria:Q9F740

```

seq_documentation_block:
ID Q9F740 PRELIMINARY; PRT; 426 AA.
AC Q9F740;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE W2X.
GN W2X.
OS Shigella sonnei.
OC plasmid pInv.
OC Bacteria: Proteobacteria: gamma subdivision: Enterobacteriaceae;
OX Shigella.
OX NCB1_TaxID=624;
RN SEQUENCE FROM N.A.
RC STRAIN=53G1;
RX PubMed=1092522;
RA Shepherd J., Wang L., Reeves P.R.;
RT "Comparison of the O antigen gene clusters of Escherichia coli
RT (Shigella) Sonnei and Plesiomonas shigelloides O17: Sonnei gained its
RT current plasmid borne O antigen genes from Plesiomonas shigelloides in
RT a recent event.";
RL Infect. Immun. 68:6056-6061(2000).
DR EMBL: AF285971; AAG17420.1; -.

```

```

DR InterPro: IPR002797; Polysacc_synt.
DR Pfam: PF01943; Polysacc_synt; 1.
KW Plasmid.
SQ SEQUENCE 426 AA: 49279 MW: C5E28DCC21038E5 CRC64;

```

alignment_scores:

Quality:	77.50	Length:	190
Ratio:	0.891	Gaps:	7
Percent Similarity:	45.789	Percent Identity:	22.632

alignment_block:

us-09-823-101-3/rev x Q9F740 ..

Align seg 1/1 to: Q9F740 from: 1 to: 426

```

677 TTTGTTTCTTCATTAATAGTGGCGCCCGCCACTAATATTTCTCGGCGT 628
93 PheMetPheTySerTyValSerAspSerIleSerLeuThrLeuVal 109
627 CTCCCTCATAGCTGCATATACTATATATCTCTCGCGTCCTCC 578
109 lTyrlIeSer.SerAlaTyRpheValCysIleSerSerGlyArgPheSer 125
577 CTCCCTCATAGCTGCATATACTATATATTTCTCTCGCGTCCTCC 549
126 LeuLeuGlnAlaValAlGlyArgPheArgCysGluLeuTyrlIeAsnIleTy 142
548 TTGTTCTTAATTTTGT.....TTCTACTCTCTCTTATTC 511
142 rSerThrIleTyrlIeGlyCysAsnLeuPheLeuSerLeuPheIleG 159
510 CCCCCCTTTAAATAATTAATCTCTCTTTTCATATTTGTATCTCC 461
159 lupro.....LeuTyrlTySer..... 164
460 TCCTGCCTAATAAATAAATAAACAACGCGCATGTGTTTGTATTTTC 411
165 .....AlaIleSe 167
410 TGTGTTGCTTATCCATCTCTCTGTGATATTCCTGTCTCTCTCT 361
167 rIlePheIleTySerIleSerIleLeuValPheSerSerHisLysCysA 184
360 CTCCCGCCCGCCCT.CATATAAA..... 339
184 snValProCysPheHisIleLysArgProSerIleLeuValTyrlLysAsp 200
338 .....TATCTCTGTATTAAT 322
201 PheLeuAspAlaThrProPheAlaIleLeuValLeuLeuAsnValIle 217
321 TGTGCGGCTGATATAATTTTGGGGGAAAAACTCATGTCCCGCATCGGG 272
217 uSerSerIleAspLeuPheIleLeuLysGluTyRPheserTyrlAsn... 232
271 GATAGCTCCNACANGTCACGTGCTATCTGTCAACCAACANTGTG 222
233 .....SerValAlaIleTyRglnValIleThrArgVal 243
221 GACACACAAGTTGTGCTC 204
244 AsnThrGlyLeuIleIle 249

```

seq_name: sp_invertebrate:Q9NG78

```

seq_documentation_block:
ID Q9NG78 PRELIMINARY; PRT; 791 AA.
AC Q9NG78;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE HERG-LIKE POTASSIUM CHANNEL.

```


OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Pelodierinae; Caenorhabditis.
 RN NCB1_TaxID=6239;
 RP SEQUENCE FROM N.A.
 RA Thomas J.H., Reiner D.J., Newton E.M.;
 RT "C. elegans unc-103 encodes a HERG-like potassium channel."
 RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF257518; AAF68999.1; -
 DR InterPro: IPR000636; Cation_chan_non_lig.
 DR InterPro: IPR001622; Channel_pore_K.
 DR InterPro: IPR002025; CNG_membrane.
 DR InterPro: IPR000595; CNG_binding.
 DR InterPro: IPR003967; Ery_channel.
 DR Pfam: PF00914; CNG_membrane.1.
 DR Pfam: PF00027; CNG_binding.1.
 DR PRINTS: PR01470; ERGCHANNEL.
 DR SMART: SM00100; CNGP.1.
 DR PROSITE: PS50042; CNGP_BINDING_3; 1.
 DR Ionic channel.
 KM SEQUENCE 791 AA; 88518 MW; 5BF65AF56C380BEA CRC64;

alignment_scores:
 Quality: 77.50 Length: 250
 Ratio: 0.760 Gaps: 10
 Percent Similarity: 40.800 Percent Identity: 20.400

alignment_block:
 US-09-823-101-3 x Q9NG78 ..

Align seg 1/1 to: Q9NG78 from: 1 to: 791

```

10 TCCTGTCACCTGGGTCCTGCTGCGACCTATGCCCTGTACACAGTGTGC 59
   |||::: ||| ||| |||::: |||
585 SerAlaGluSerAlaSerArgCysAspSerAsnProIleAspArgArg1 601
60 TTGCGCTGGATCTTCTTGAGTCT .....C 85
   ||| ||| ||| ||| |||
601 nSerAlaGlySerArgSerSerArgCysSerProProHISAlaLal 618
86 TGCTATGATCTCTGTCCTAGCGTATCTATATTT .....CTCAAC 126
   ||::: ||| ||| |||::: |||
618 eunHraLarThrArgSerGluAlaThrProLeuAlaArgArgSerThrAsn 634
127 AGTCATATGATGACACAGACAGCAAGTGGCGTGAACGTGGCATGACGT 176
   |||::: ||| ||| ||| |||
635 HIsHISgluGluAsp .....AspAlaLe 642
177 TCACGATGACGTGTCACCAACATGAGACACACTGTGTGTCCACAN 226
   : ||| |||:::
642 uPheAspAspIle ..... 646
227 TGGTTTGAACAGATATAGCACACACGTGACNTGTGGAGCTATCCCGA 276
   ::::: |||::: |||
647 .....ArgAlaPheAlaArgGlyAsnThrValThrMetSerProThr 660
277 TGGGGGACATGAGTTTTCCTCCCAAAATTTATCCACCCACCAATTA 326
   ::: ||| ||| ||| ||| |||
661 ValAlaGlyAsnSerValSerPro .....ThrThrAlaIleH1 673
327 TACACAGAGATATTTATATAGAGGGGGGGGAGAGACAGCGG .....A 370
   : ||| ||| ||| ||| |||
673 sAsn .....AspGlyIleHiss 679
371 CACAGGAATATCAACAAG ..... 390
679 erGInGluLeuSerAspArgSerAspArgGluGluArgAlaAsn 695
391 .....AAGATGATTAACGAAA 407
696 MetPheGlyArgArgLeuGluSerIleGluSerGluMetGluArgMetG1 712

```

```

408 ACAGAAATACAAAACACATGCGCGTATTTTATTTATAGGCA 457
   ::|||::: |||::: |||::: |||:::
712 nAsnLysPheAsnSerAspMetGluThrLeuIleLysLeuVal ..... 726
458 GGAGAGATTACAAATATGAAAAGACAGATTAATTTTAAAGGG 507
   ::|||::: |||::: |||::: |||:::
727 .....LysGluGlnSerIleIleArgAsnAsn 735
508 GGGGATTAAGAGAGAGTGAACAAAATTAAGACAAAACATATT 557
   |||::: |||::: |||::: |||:::
736 GlySerSerAsnGluGluProAsnAlaArgTyrArgProAsnAsnYrt11 752
558 TTTT .....GTTGGGGGGGAGAGAGAGACACAGCAG 589
   : ||| ||| ||| ||| ||| ||| |||
752 eSerSerAlaIleArgLeuProAsnGlyGlyGlyAlaValAlaSPG 769
590 AGAGAAATTAATTAATATGATGACGCTATGAGAGAGACCCGAGAGAA 639
   ||::: ||| ||| ||| ||| ||| |||
769 IuMetArgValSerArgLeuSerSerHISgluProThrProThrGln 785

```

seq_name: sp_invertebrate:062186
 seq_documentation_block:
 ID 062186 PRELIMINARY; PRT; 384 AA.
 AC 062186;
 DT 01-AUG-1998 (TREMblrel. 07, Created)
 DT 01-AUG-1998 (TREMblrel. 07, Last sequence update)
 DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
 DE F28D9.2 PROTEIN.
 GN F28D9.2.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Pelodierinae; Caenorhabditis.
 RN NCB1_TaxID=6239;
 RP SEQUENCE FROM N.A.
 RA Dobson R.;
 RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94150718; PubMed=7906398;
 RA Wilson R., Alnsough R., Anderson K., Baynes C., Berks M.,
 RA Bonfield J., Burton J., Connell M., Copey T., Cooper J., Coulson A.,
 RA Craxton M., Dear S., Du Z., Durbin R., Favella A., Fulton L.,
 RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
 RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
 RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
 RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,
 RA Smaison N., Smith A., Sonnenhammer E., Staden R., Sulston J.,
 RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
 RA Watson A., Weinstock L., Wilkinson-Sproat J., Woldman P.,
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 RT elegans."
 RL Nature 368:32-38(1994).
 DR EMBL: Z81518; CAB04215.1; -
 DR InterPro: IPR000168; 7TM_nematode.
 DR InterPro: IPR003003; 7TM_chemorecept_2.
 DR Pfam: PF01604; 7tm_5; 1.
 SO SEQUENCE 384 AA; 44832 MW; EE96FDP9DAB9F581C CRC64;

alignment_scores:
 Quality: 77.00 Length: 204
 Ratio: 0.837 Gaps: 9
 Percent Similarity: 45.098 Percent Identity: 22.549

alignment_block:
 US-09-823-101-3/rev x 062186 ..

Align seg 1/1 to: 062186 from: 1 to: 384

```

524 CTCTCTCTTATTTCCCGCCCGC...TTTAAATAATAATCTCTCT..... 483

```

```

      :: |||||::|||:: ||| |||::
7 IleCysProTyrGlyProIleAspTyrTyrMetLysThrLeuGlnLeu1 23
482 .....TCTTCATATTGTGATTC 464
23 eGlyLeuPheSerLeuProIleAsnPheLeuCysPheTyrPheValTyrP 40
463 TCCTCCGCGCTAATAAAAAAATAACACCGCATGCTGTTTGTAT 414
   |||::|||::|||:: ||| |||:: |||
40 heTyrThrProLysGlySerLysPhe.....ArgTyrCysLeuAlaTyr 54
413 TTCGTTTTCGTTATCCATCTTCCTGTTGATTTCTGTC...CC 367
   ||| ||| ||| ||| ||| ||| ||| ||| |||
55 PheGlnPheIleAlaPheLeuValGlnValAspMetSerLeuValCysPr 71
366 GTCTCTGCCCCCCCCCTCATATAAATATCTGTGTTATATTGCTG 317
   |||::|||::|||:: ||| |||:: ||| ||| |||
71 oGlyPhe.....TyrLeuPheProLeuMetGlyG 81
316 GGGTGGATAAATTTTGGGGGAAACATCATGTCCTCCCATCGGGGATAG 267
   |||::|||::|||:: ||| |||:: ||| |||
81 LyTyrAsnLeuAlaGluThrAsnArgLeuPheSerGlyHisGln..... 95
266 CTCGCACACAGTCACGTTGCTCTATCTCTGTTCAACACANTGGGACAC 217
   |||::|||::|||:: ||| |||:: ||| |||
96 .....ThrValAlaPheGlyTyrPheLysPheSerPheGluLe 108
216 ACAAGTTGCTCTCATGTTGGTGGACACAGTCATCGAGACTGCA.... 171
   |||::|||::|||:: ||| |||:: ||| |||
108 uProSerLeuLeuLeuCysPheIleTyrArgHisAsnAlaAlaAsnPr 125
171 ..... 171
125 heAsnProLysPheLysIleArgIleSerIleGlnTyrLeuValIleAla 141
170 ...TGCCACACAGTCCACGCACTGCTGT..... 144
142 ThrCysHisThrPheProPheValThrAlaIleCysLeuLeuLysSerG 158
   |||::|||::|||:: ||| |||:: ||| |||
143 .CTGTCATGCATGACTGTTGAGATTTATGATAGCTAGG..... 102
158 uLeuSerHisValGlnGlnValArgIleLeuGlnLysValArgArgTyrL 175
101 .....ACCAGATGCATAGCAGAGACTCACAGAAGATCCAGGCAAA 60
   |||::|||::|||:: ||| |||:: ||| |||
175 LysLeuGlnThrArgGluThrAlaAsnTyrThrGlnAsnThrProAsnCy 191
59 GCACACCTGGTG 48
   |||::|||::|||:: ||| |||:: ||| |||
192 LeuHisIleLeu 195

seq_name: sp_fungi:Q9P472

seq_documentation_block:
ID Q9P472 PRELIMINARY: PRT; 472 AA.
AC Q9P472;
DT 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE CAMP-DEPENDENT PROTEIN KINASE PKAC CATALYTIC SUBUNIT.
GN PKAC.
OS Emerizella nidulans (Aspergillus nidulans).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiiales; Trichocomaceae; Emerizella.
OX NCBI_TaxID=5072;
RN [1]
RP SEQUENCE FROM N.A.
RT Shimizu K., Keller N.P.;
RT "Cloning of PKAC, a gene encoding a CAMP-dependent protein kinase
RT catalytic subunit of Aspergillus nidulans."
RT Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
CC -1. SIMILARITY: TO THE SER/THR FAMILY OF PROTEIN KINASES.
DR EMBL; AF262987; AAF57562.1; -

```

```

DR InterPro: IPR000719; Euk_pkinase.
DR InterPro: IPR000961; Pkinase_C.
DR InterPro: IPR002290; Ser_thr_kin_actsite.
DR InterPro: IPR001245; Tyr_kin.
DR Pfam: PF00069; Pkinase_1.
DR Pfam: PF00433; Pkinase_C_1.
DR PRINTS: PR00109; TYRKINASE.
DR SMART: SM00220; S_TKc_1.
DR SMART: SM00133; S_TK_X_1.
DR PROSITE: PS00107; PROTEIN_KINASE_APP; 1.
DR PROSITE: PS00111; PROTEIN_KINASE_DOM; 1.
DR PROSITE: PS00108; PROTEIN_KINASE_ST; 1.
DR AMP-binding: kinase; Serine/threonine-protein kinase; Transferase.
SQ SEQUENCE 472 AA; 53233 MW; D3187698050AD8A CRC64;

```

```

alignment_scores:
Quality: 77.00 Length: 139
Ratio: 1.305 Gaps: 7
Percent Similarity: 42.446 Percent Identity: 25.899

```

```

alignment_block:
US-09-823-101-3 x Q9P472 ..

```

```

Align seg 1/1 to: Q9P472 from: 1 to: 472

```

```

26 CTGCTGGTGGACCTATGCGTGTGCACACAGGTGCTGTCCTGATCTTCT 75
   :: |||||::|||:: ||| |||:: |||:: |||::
312 lIeHrTrPrHrLeuCysGlyThrProAspTyrLeuAlaProGluValA 328
76 TGTGAGTCCTGCTATGATCC.....TGGTCTAGCGTATCC 113
   ::||| ||| ::||| ||||| |||||
328 lAlaSerLysGlyTyrAsnLysSerValAspTrpTrpSerLeuLyllel 345
114 TATATTCTCACAGCTCATATCGATGACAGACAGCAAGTCCGTGGAAC 163
   || |||:: ||| ::||| |||:: ||| |||
345 euIlePheGluMetLeu..... 350
164 TGTGCGATGACGTTACAGATGACGTGTCACCAACAATGAGAGCAACT 213
   ||||| ||||| |||||
351 CysGly.....PheThr..... 354
214 TGTGTCTCACANTGTTTGACAGACT.....ATAGCACACAGCTG 254
   ||| ||| ::||| ||| |||:: |||
355 .....ProheTrpAspGlnGlySerProValLysIleTyrGlnsni 369
255 ACNTGTGNGAGCTATCCCGATGGGGGACATGATTTTCCCCCAAAA 304
   ::||| ||| ::||| ||| ||||| |||||
369 lLeuAlaGlyArgIleLys.....PheProProLys 379
305 ATTATCCACCC.....CACCAATTATAACCAAGATAT 339
   ::||| ||| ::||| ||| ||||| |||||
380 LeuHisProAspAlaValAspLeuLeuSerArgLeuLeuIleTrpSerSple 396
340 TTATATAGGGGGGGGAGAGACAGAGAGAGCAAGAAATATCAACAG 389
   : :: |||:: ||| ::||| ||| ::|||
396 uThrLysArgLeuLysAsnLeuHisGlyLysProAspAspLysAsnH 413
390 GAAGATGATTAACGAA 406
   |||::|||::|||:: ||| |||:: ||| |||
413 lSProTrpPheAlaGlu 418

```

THIS PAGE BLANK (USPTO)

OM of: US-09-823-101-4 to: A_Geneseq_1101.* out_format : pfs

Date: Jan 17, 2002 3:59 PM

About: Results were produced by the GenCore software, version 4.5.
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlp
-O=/cgnr2_1/USPRO.spool/US09823101/runtat_17012002_155137_22003/app_query.fasta_1.2239
-DB=A_Geneseq_1101 -OFMT=fasta -SUFFIX=n2p.rag -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOPEXT=0.000
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blotsum62
-TRANS=human40.cdi -LIST=45 -DOCLIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext MINLEN=0 -MAXLEN=2000000000
-USER=US09823101_GCGN1_1195 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPEXY
-WAIT -THREADS=1

Search information block:

Query: US-09-823-101-4
Query length: 509
Database: A_Geneseq_1101.*
Database sequences: 522463
Database length: 74073290
Search time (sec): 146.120000

score_list:

Sequence	Strid	Orig	ZScore	EScore	Len	Documentation	..	454	1
/SID58/gcgdata/geneseq/geneseqp/AA2000.DAT:AAAB41897					101.50	186.21	0.0047	454	1
/SID58/gcgdata/geneseq/geneseqp/AA2001.DAT:AAAB40851					99.00	170.65	0.0119	1310	1
/SID58/gcgdata/geneseq/geneseqp/AA1999.DAT:AAV122206					97.00	179.24	0.0144	361	1
/SID58/gcgdata/geneseq/geneseqp/AA2000.DAT:AAAG23678					96.50	181.22	0.0152	266	1
/SID58/gcgdata/geneseq/geneseqp/AA2000.DAT:AAAG23677					96.50	181.04	0.0152	271	1
/SID58/gcgdata/geneseq/geneseqp/AA1995.DAT:AAAR79912					96.50	171.60	0.0196	707	1
/SID58/gcgdata/geneseq/geneseqp/AA1999.DAT:AAAB84052					96.50	171.60	0.0196	707	1
/SID58/gcgdata/geneseq/geneseqp/AA2001.DAT:AAAB48964					96.00	169.62	0.0229	779	1
/SID58/gcgdata/geneseq/geneseqp/AA2000.DAT:AAAB42438					96.00	169.62	0.0229	779	1
/SID58/gcgdata/geneseq/geneseqp/AA2000.DAT:AAAB68784					96.00	165.91	0.0253	1115	1
/SID58/gcgdata/geneseq/geneseqp/AA2001.DAT:AAAE04368					96.00	165.27	0.0257	1212	1
/SID58/gcgdata/geneseq/geneseqp/AA1999.DAT:AAV55954					96.00	165.10	0.0259	1233	1
/SID58/gcgdata/geneseq/geneseqp/AA1999.DAT:AAV55931					96.00	165.05	0.0259	1239	1
/SID58/gcgdata/geneseq/geneseqp/AA2000.DAT:AAAB43016					95.50	164.63	0.0281	1165	1
/SID58/gcgdata/geneseq/geneseqp/AA1994.DAT:AAAB56494					94.00	166.06	0.0382	737	1
/SID58/gcgdata/geneseq/geneseqp/AA1997.DAT:AAAM06084					94.00	166.06	0.0382	737	1
/SID58/gcgdata/geneseq/geneseqp/AA1997.DAT:AAAM25019					94.00	166.06	0.0382	737	1
/SID58/gcgdata/geneseq/geneseqp/AA2001.DAT:AAAB39441					93.50	173.16	0.0351	323	1
/SID58/gcgdata/geneseq/geneseqp/AA2001.DAT:AAAG75411					93.50	172.45	0.0358	347	1
/SID58/gcgdata/geneseq/geneseqp/AA2001.DAT:AAAG75411					92.50	165.11	0.0556	594	1
/SID58/gcgdata/geneseq/geneseqp/AA2001.DAT:AAAM40973					92.00	168.64	0.0541	374	1
/SID58/gcgdata/geneseq/geneseqp/AA2000.DAT:AAAB43954					92.00	168.59	0.0542	376	1
/SID58/gcgdata/geneseq/geneseqp/AA2001.DAT:AAAB39187					92.00	168.91	0.0567	446	1
/SID58/gcgdata/geneseq/geneseqp/AA2001.DAT:AAAB67047					92.00	161.96	0.0647	737	1
/SID58/gcgdata/geneseq/geneseqp/AA1999.DAT:AAV29039					91.50	162.78	0.0702	611	1
/SID58/gcgdata/geneseq/geneseqp/AA2001.DAT:AAAB36624					91.00	172.04	0.0609	215	1
/SID58/gcgdata/geneseq/geneseqp/AA2000.DAT:AAAG15789					91.00	171.38	0.0619	230	1
/SID58/gcgdata/geneseq/geneseqp/AA2000.DAT:AAAG52491					91.00	171.38	0.0619	230	1
/SID58/gcgdata/geneseq/geneseqp/AA2001.DAT:AAAB94010					91.00	159.98	0.0840	732	1
/SID58/gcgdata/geneseq/geneseqp/AA2000.DAT:AAAG20196					89.50	170.83	0.0859	178	1
/SID58/gcgdata/geneseq/geneseqp/AA2000.DAT:AAAG49267					89.50	170.83	0.0859	178	1
/SID58/gcgdata/geneseq/geneseqp/AA2000.DAT:AAAG20195					89.50	164.96	0.1005	323	1
/SID58/gcgdata/geneseq/geneseqp/AA2000.DAT:AAAG49266					89.50	164.96	0.1005	323	1
/SID58/gcgdata/geneseq/geneseqp/AA2000.DAT:AAAG49265					89.50	164.69	0.1012	332	1
/SID58/gcgdata/geneseq/geneseqp/AA2001.DAT:AAAB93238					89.00	155.03	0.1454	798	1
/SID58/gcgdata/geneseq/geneseqp/AA1998.DAT:AAAB53571					89.00	153.04	0.1533	976	1
/SID58/gcgdata/geneseq/geneseqp/AA1998.DAT:AAAB41377					89.00	153.04	0.1533	976	1
/SID58/gcgdata/geneseq/geneseqp/AA2001.DAT:AAAM40294					88.00	151.43	0.1600	1150	1
/SID58/gcgdata/geneseq/geneseqp/AA2001.DAT:AAU03590					88.50	161.00	0.1375	392	1
/SID58/gcgdata/geneseq/geneseqp/AA2001.DAT:AAAB92959					88.50	161.00	0.1375	392	1

/SID58/gcgdata/geneseq/geneseqp/AA2001.DAT:AAAB93038	88.50	161.00	0.1375	392
/SID58/gcgdata/geneseq/geneseqp/AA2001.DAT:AAAB94332	88.50	161.00	0.1375	392
/SID58/gcgdata/geneseq/geneseqp/AA1999.DAT:AAV40999	88.50	156.73	0.1542	605
/SID58/gcgdata/geneseq/geneseqp/AA2000.DAT:AAAB18188	88.50	156.08	0.1569	646
/SID58/gcgdata/geneseq/geneseqp/AA2001.DAT:AAAB99675	88.50	149.61	0.1864	1246

seq_name: /SID58/gcgdata/geneseq/geneseqp/AA2000.DAT:AAAB41897

seq_documentation_block:

ID AAB41897 standard; Protein: 454 AA.
XX
AC AAB41897;
XX
DT 08-FEB-2001 (first entry)
XX
DE Human ORFX ORF1661 polypeptide sequence SEQ ID NO:3322.
XX
KW Human; open reading frame; ORFX; detection; cytosolic; hepatotropic;
KW vulnery; antipositive; antiparkinsonian; noctropic; neuroprotective;
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiac;
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antineumatic; antihypertensive;
KW antiviral; antibacterial; antifungal; antineumatic; antihypertensive;
KW antineumatic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antineumatic disease; coagulation;
KW thrombosis; contraceptive.
XX
OS Homo sapiens.
XX
PN WO200058473-A2.
XX
PD 05-OCT-2000.
XX
PE 31-MAR-2000; 2000WO-US08621.
XX
PR 31-MAR-1999; 98US-0127607.
PR 02-APR-1999; 98US-0127636.
PR 05-APR-1999; 99US-0127728.
PR 30-MAR-2000; 2000US-0540763.
XX
PA (CUBA-) CUBAGEN CORP.
XX
PI Shinkets RA, Leach M;
XX
DR WPI; 2000-602362/57.
XX
N-PSDB; AAC76106.
XX
PT Novel nucleic acids and peptides derived from open reading frame X,
PT useful for treating e.g. cancers, proliferative disorders,
PT neurodegenerative disorders and cardiovascular disease -
XX
PS
XX
Claim 11; Page 2511-2512; 5507pp; English.
CC AAC4446 to AAC7606 encode the proteins given in AAB40237 to AAB43397,
CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC sequences have activities such as: cytosolic; hepatotropic; vulnery;
CC antipositive; antiparkinsonian; noctropic; neuroprotective;
CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
CC immunostimulant; cardiac; thrombolytic; coagulant; vasotropic;
CC antidiabetic; hypotensive; dermatological; immunosuppressive;
CC antineumatic; antibacterial; antifungal; antineumatic; antihypertensive;
CC antihypertensive; antineumatic. The sequences can be used for determining
CC the presence of or predisposition to, or preventing or treating
CC pathological conditions associated with an ORFX-associated disorder. The
CC nucleic acids can be used to express ORFX proteins in gene therapy
CC vectors. The proteins and nucleic acids may be used to treat cancers,
CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC graft vs host disease, cardiovascular disease, diabetes mellitus,

CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
CC erythematous, severe combined immunodeficiency (SCID), AIDS, viral,
CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
CC coagulation; to inhibit thrombosis; and as a contraceptive.

XX Sequence 454 AA;

Alignment_scores:

Quality: 101.50 Length: 162
Ratio: 0.976 Gaps: 5
Percent Similarity: 64.198 Percent Identity: 24.691

Alignment_block:

US-09-823-101-4 x AAM41897 ..

Align seg 1/1 to: AAM41897 from: 1 to: 454

```
30 ATCAGCGAAGAAACGTAATCTATCA.....GAGGAAAAAANA 70
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
262 ValArgGluAlaGlnSerPheIleSerAlaAlaIleGluProGlnSerG1 278
71 AAAAAAANAACCCGGGGGAAACGAGGACATACGAGTG.. 118
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
278 yLsSerAsnGluArgLysGlyLysArgSerGlnSerHisThrArgSerL 295
119 .....TCCGGGCTGGAACACTGCGTACCCGGGTACAAAC 154
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
295 ySerArgSerSerLysSerHisSerArgArgLysArgSerGlnSer 311
155 TCCAAAGAGAAACATGTCACACGCTCCACCTACCGCGACACAGCTA 204
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
312 LysHisArgSerArgSerHisArgSerArgSerArgGlnLysAspArg 328
205 GAGAAACACATCTACTAGAACCTCAACATATATATACATAGTACAGAA 254
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
328 GArgArgSerLysSerProHisLysLysArgSerLysSerArgGluArg 345
255 ACGCCACATAGAGCTACCGCATTAAGTACAACTCGCAAGACACGCTAT 304
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
345 rGlySerArgSerArgSerHisSerArgAspLysArgLysAspThrArg 361
305 CAGAGACTGACGAGAGAGAAACGAGAGAGAGAGACAAACACAGAGC 354
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
362 GluLysIle.LysGluLysGluArgValLysGluLysAspArgLys. 377
355 CAGCGTGAGCGCGATGAGGAGAGAGGAGATACCCGAAAGAAACGGGAA 404
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
378 GluArgGluArg...GluLysGluArgGluLysGluLysGluArgGly 393
405 AAGACAGATGAAGTATCTCAGTTGACGAAGACCAAAAGCAGAGAGC 454
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
393 SasnLysAspArgAspLysGluArgGluLysAspArgGluLysAspLysG 410
455 ATGTTGATCAAGAAACGAGCGAGACAGAC 486
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
410 LuLysAspArgGluArgGluArgGluLysGlu 420
```

seq_name: /std58/gcdata/geneseq/geneseq/AA2001.DAT:AAM40851

seq_documentation_block:

ID AAM40851 standard; protein; 1310 AA.

XX AAM40851;

XX 22-OCT-2001 (First entry)

XX Human polypeptide SEQ ID NO 5782.

XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;

KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia.

OS Homo sapiens.

XX WO200153312-A1.

XX 26-JUL-2001.

XX 26-DEC-2000; 2000WO-US34263.

XX 21-JAN-2000; 2000US-0488725.

XX 25-APR-2000; 2000US-0552317.

XX 09-JUL-2000; 2000US-0598042.

XX 19-JUL-2000; 2000US-0620312.

XX 03-AUG-2000; 2000US-0653450.

XX 14-SEP-2000; 2000US-0662191.

XX 19-OCT-2000; 2000US-0693036.

XX 29-NOV-2000; 2000US-0727344.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;

XX WPI; 2001-442253/47.

XX N-PSDB; AAI60007.

XX Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -

PS Example 2; SEQ ID NO 5782; 10078pp; English.

XX The invention relates to human nucleic acids (AA157798-AA161369) and
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.

CC Note: The sequence data for this patent did not form part of the printed
CC specification.

XX Sequence 1310 AA;

Alignment_scores:

Quality: 99.00 Length: 158
Ratio: 1.138 Gaps: 5
Percent Similarity: 55.063 Percent Identity: 25.949

Alignment_block:

US-09-823-101-4 x AAM40851 ..

Align seg 1/1 to: AAM40851 from: 1 to: 1310

```
28 TCATCAGGAAAAAAGCTAAT.....TCTATCAGAGAGAA 62
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
666 SerSerGlyGluLysGlyGlyTyrGluIlePheValLysAspThr 682
63 AAAAAAANAACCCGGGGGAAACGAGGACATAG 112
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
682 rHisGluLysSerLysAsnLysAsnArgAspLys..... 694
```

[illegible]

alignment_scores:	
Quality:	97.00
Ratio:	1.032
Percent Similarity:	59.494
	Length:
	Gaps:
	5
Percent Identity:	22.152

alignment_block:
US-09-823-101-4 x AAY22206

Align seg 1/1 to: AAY22206 from: 1 to: 361

[illegible]

PR	25-FEB-1999;	99US-0121825.
PR	05-MAR-1999;	99US-0123180.
PR	09-MAR-1999;	99US-012348.
PR	23-MAR-1999;	99US-0125788.
PR	25-MAR-1999;	99US-0126264.
PR	29-MAR-1999;	99US-0126785.
PR	01-APR-1999;	99US-0127462.
PR	06-APR-1999;	99US-0128334.
PR	08-APR-1999;	99US-0128714.
PR	16-APR-1999;	99US-0129845.
PR	19-APR-1999;	99US-0130077.
PR	21-APR-1999;	99US-0130409.
PR	23-APR-1999;	99US-0130510.
PR	28-APR-1999;	99US-0130891.
PR	30-APR-1999;	99US-0131449.
PR	30-APR-1999;	99US-0132048.
PR	04-MAY-1999;	99US-01332407.
PR	05-MAY-1999;	99US-01332484.
PR	06-MAY-1999;	99US-0133485.
PR	06-MAY-1999;	99US-0133486.
PR	07-MAY-1999;	99US-0133487.
PR	11-MAY-1999;	99US-0133863.
PR	14-MAY-1999;	99US-0134256.
PR	14-MAY-1999;	99US-0134218.
PR	14-MAY-1999;	99US-0134319.
PR	14-MAY-1999;	99US-0134321.
PR	18-MAY-1999;	99US-0134370.
PR	19-MAY-1999;	99US-0134768.
PR	20-MAY-1999;	99US-0134961.
PR	21-MAY-1999;	99US-0135124.
PR	21-MAY-1999;	99US-0135353.
PR	25-MAY-1999;	99US-0135629.
PR	27-MAY-1999;	99US-0136021.
PR	28-MAY-1999;	99US-0136392.
PR	01-JUN-1999;	99US-0137282.
PR	03-JUN-1999;	99US-0137222.
PR	04-JUN-1999;	99US-0137528.
PR	07-JUN-1999;	99US-0137502.
PR	08-JUN-1999;	99US-0137724.
PR	10-JUN-1999;	99US-0138094.
PR	10-JUN-1999;	99US-0138540.
PR	10-JUN-1999;	99US-0138847.
PR	16-JUN-1999;	99US-0139119.
PR	16-JUN-1999;	99US-0139452.
PR	17-JUN-1999;	99US-0139457.
PR	18-JUN-1999;	99US-0139492.
PR	18-JUN-1999;	99US-0139454.
PR	18-JUN-1999;	99US-0139455.
PR	18-JUN-1999;	99US-0139456.
PR	18-JUN-1999;	99US-0139457.
PR	18-JUN-1999;	99US-0139458.
PR	18-JUN-1999;	99US-0139459.
PR	18-JUN-1999;	99US-0139459.
PR	18-JUN-1999;	99US-0139460.
PR	18-JUN-1999;	99US-0139461.
PR	18-JUN-1999;	99US-0139462.
PR	18-JUN-1999;	99US-0139463.
PR	18-JUN-1999;	99US-0139750.
PR	21-JUN-1999;	99US-0139763.
PR	22-JUN-1999;	99US-0139817.
PR	23-JUN-1999;	99US-0139839.
PR	24-JUN-1999;	99US-0140354.
PR	24-JUN-1999;	99US-0140355.
PR	28-JUN-1999;	99US-0140659.
PR	29-JUN-1999;	99US-0140991.
PR	30-JUN-1999;	99US-0141281.
PR	01-JUL-1999;	99US-0141842.
PR	01-JUL-1999;	99US-0142154.
PR	02-JUL-1999;	99US-0142055.
PR	06-JUL-1999;	99US-0142360.
PR	08-JUL-1999;	99US-0142803.
PR	09-JUL-1999;	99US-0142920.
PR	12-JUL-1999;	99US-0142977.
PR	13-JUL-1999;	99US-0143542.
PR	14-JUL-1999;	99US-0143624.
PR	15-JUL-1999;	99US-0144005.
PR	16-JUL-1999;	99US-0144085.
PR	16-JUL-1999;	99US-0144086.
PR	19-JUL-1999;	99US-0144325.
PR	19-JUL-1999;	99US-0144331.
PR	19-JUL-1999;	99US-0144332.
PR	19-JUL-1999;	99US-0144333.
PR	19-JUL-1999;	99US-0144334.
PR	19-JUL-1999;	99US-0144335.
PR	20-JUL-1999;	99US-0144352.
PR	20-JUL-1999;	99US-0144632.
PR	21-JUL-1999;	99US-0144884.
PR	21-JUL-1999;	99US-0145086.
PR	21-JUL-1999;	99US-0145088.
PR	22-JUL-1999;	99US-0145085.
PR	22-JUL-1999;	99US-0145087.
PR	22-JUL-1999;	99US-0145089.
PR	22-JUL-1999;	99US-0145192.
PR	22-JUL-1999;	99US-0145145.
PR	23-JUL-1999;	99US-0145218.
PR	24-JUL-1999;	99US-0145224.
PR	26-JUL-1999;	99US-0145276.
PR	27-JUL-1999;	99US-0145913.
PR	27-JUL-1999;	99US-0145918.
PR	27-JUL-1999;	99US-0145919.
PR	28-JUL-1999;	99US-0145951.
PR	02-AUG-1999;	99US-0146388.

PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 21-JUN-1999; 99US-0139763.
PR 22-JUN-1999; 99US-0139817.
PR 23-JUN-1999; 99US-0139859.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 04-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148584.

PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149724.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151433.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157113.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 08-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

alignment_scores:

Quality: 96.50 Length: 128
Ratio: 1.664 Gaps: 3
Percent Similarity: 45.312 Percent Identity: 22.656

alignment_block:

US-09-823-101-4 x AAG23677

Align seg 1/1 to: AAG23677 from: 1 to: 271

28 TCATCAGGAGAAAAACGTAATTCATCAAGAGAAAAA

FT Modified-site /note= "potential phosphorylation site"
FT 543
FT Modified-site /note= "potential phosphorylation site"
FT 550
FT Modified-site /note= "potential phosphorylation site"
FT 554
FT Modified-site /note= "potential phosphorylation site"
FT 570
FT Modified-site /note= "potential glycosylation site"
FT 572
FT Modified-site /note= "potential phosphorylation site"
FT 624
FT Modified-site /note= "potential phosphorylation site"
FT 625
FT Modified-site /note= "potential phosphorylation site"
FT 632
FT Modified-site /note= "potential phosphorylation site"
FT 681
FT Modified-site /note= "potential phosphorylation site"
FT 682
FT Modified-site /note= "potential phosphorylation site"
FT 688
FT Modified-site /note= "potential phosphorylation site"
FT 689
FT Modified-site /note= "potential phosphorylation site"
FT 706
FT Modified-site /note= "potential phosphorylation site"
FT 718
FT Modified-site /note= "potential glycosylation site"
FT 720
FT Modified-site /note= "potential phosphorylation site"
FT 726
FT Modified-site /note= "potential phosphorylation site"
FT 811
FT Modified-site /note= "potential phosphorylation site"
FT 815
FT Modified-site /note= "potential phosphorylation site"
FT 836..1115
FT Domain /note= "NIKI-like kinase domain"
FT 898
FT Modified-site /note= "potential phosphorylation site"
FT 931
FT Modified-site /note= "potential phosphorylation site"
FT 958
FT Modified-site /note= "potential phosphorylation site"
FT 978
FT Modified-site /note= "potential phosphorylation site"
FT 999
FT Modified-site /note= "potential phosphorylation site"
FT 1012
FT Modified-site /note= "potential phosphorylation site"
FT 1067
FT Modified-site /note= "potential glycosylation site"
FT 1113
FT Modified-site /note= "potential phosphorylation site"
XX
XX MO20006728-A2.
XX
XX 10-FEB-2000.
XX
XX 28-JUL-1999;
XX 99WO-US17132.
XX
XX 28-JUL-1998;
XX 98US-0123494.
XX 14-SEP-1998;
XX 98US-0152814.
XX 14-OCT-1998;
XX 98US-0173482.
XX 03-NOV-1998;
XX 98US-0106889.
XX 19-NOV-1998;
XX 98US-0109093.
XX 22-DEC-1998;
XX 98US-0113796.
XX 12-JAN-1999;
XX 99US-0173482.
XX 12-JAN-1999;
XX 99US-0229005.
XX
XX (INCYTE) INCYTE PHARM INC.

PI Hillman JL, Lai P, Tang YT, Corley NC, Guegler KJ, Baughn MR;
PI Patterson C, Bandman O, Au-Young J, Gorgone GA, Yue H, Azimzai Y;
PI Reddy R, Lu DAM, Shih LL,
XX
XX WPI: 2000-183125/16.
DR N-PSDB; AA246153.
XX
XX New human phosphorylation effectors useful for the diagnosis, treatment
XX and prevention of proliferative, immune and neuronal disorders
XX
XX Claim 1: Page 98-100; 142pp; English.
PS
XX
XX AAY68769-95 and AAY68797-99 represent human phosphorylation effectors
XX (PSP), designated PHS1-PHS31 (the protein sequence for PHS28 is not
XX given in the specification). The sequences were isolated from cDNA
XX libraries prepared from various human tissues. The PSP proteins are
XX useful for the diagnosis, treatment and prevention of proliferative
XX disorders, immune disorders and neuronal disorders. The PSP proteins
XX form pharmaceutical compositions which useful for treating or preventing
XX disorders associated with decreased PSP expression/activity. PSP
XX antagonists are useful for treating or preventing disorders associated
XX with increased PSP expression/activity.
SQ Sequence 1135 AA;

alignment_scores: length: 174
 Quality: 96.00 Gaps: 5
 Ratio: 1.043 Percent Identity: 21.839
Percent Similarity: 52.874

alignment_block:
US-09-823-101-4 x AAY68784

Align seg 1/1 to: AAY68784 from: 1 to: 1135

43 CGTAATTCATCAAGAGAAAAA..... 84
 ||||| ||||| |||||
303 LysAspHisIleAspArgThrArgLysArgGlyGluLysAspLunh 319
85AAC 88

319 rgluYrgluTySerGlySerGluGluGluGluGluValProGluG 336
89 CCGGGGGAACCGAGCATAGCGATGCCGGCGTGA...CTG 135
 ||||| ||||| ||||| |||||
336 lngluGlyGluProSerSerIleValAsnValProGlyGluSerThrLeu 352
136 TGCTACCGGCTCAAACTCCAAAGAGAAACATTGTCAACGCCCTCC 185
 ||||| ||||| ||||| |||||
186 ACTACCGCGCACACGCTAGAGAAACATCTACTGAAACTCACACAT 235
369 AleuArgArgGlnGlnLeuLeuGln.....GlnG 380

236 ATATACATAGTACAGAAAGCCACATAGACTACGCA..... 276
380 lngluArgGluGlnGluGluTyTyLysArgGlnLeuLeuValGluGln 396
277TAAATCAACTCGCAAGACAGCATTCACAGACTGA 314
397 LysArgIleGluGlnGlnLysGlnGlnArgArgArgLeuGluGln 413
315 CGAGAGAGAAACGAGACAGAGAGACAAACACAGACCGCTGAGC 364
 ||||| ||||| ||||| |||||
413 nArgArgGluArgGluAlaArgGlnGlnGlnGluGlnGlnArgA 430
365 GCGATGAGCAGCGCAGTACCGAAAGAAAGGGGAAAGCAGGANG 414
430 rglGlnGlnGluGluLysArgArgLeuGlnGlnGlnGluGlnArgArgLys 446
415 AAATATCTCAGTTGACGAGACCAAAAGCAGAGAGATGTGATCA 464

KW ischemic disorder; inflammation; diabetes mellitus; fibrosis; mitosis;
 KW mesangial disorder; growth regulation; wound healing; T cell activation;
 KW immunosuppressant.
 XX
 OS Homo sapiens.
 XX
 PN WO995036-A2.
 XX
 PD 21-OCT-1999.
 XX
 PF 13-APR-1999; 99WO-US08150.
 XX
 PR 14-APR-1998; 98US-0081784.
 XX
 PA (SUGEN-) SUGEN INC.
 XX
 PI Plowman G, Martinez R, Whyte D;
 XX
 DR WPI; 1999-611301/52.
 DR N-PSDB; AA240483.
 XX
 PT Novel kinase-related polypeptides used for the diagnosis and treatment
 of kinase-related diseases and disorders -
 PS
 XX Claim 11; Page 269-274; 387pp; English.
 XX
 CC This sequence represents a novel STE20-related protein kinase. The
 CC invention relates to nucleic acid molecule encoding a kinase polypeptide
 CC selected from STIK2, STIK3, STIK4, STIK5, STIK6, STIK7, ZC1, ZC2, ZC3,
 CC ZC4, KHS2, SULU1, SULU3, GER2, PAK4 and PAK5. The proteins are used to
 CC identify agonists and antagonists, and to raise antibodies. The
 CC polynucleotides are useful in gene therapy protocols. The polynucleotides,
 CC polypeptides, antibodies, antagonists and agonists may be used to treat
 CC diseases such as immune-related disorders and diseases (e.g. rheumatoid
 CC arthritis, atherosclerosis, chronic inflammatory bowel disease (e.g.
 CC Crohn's disease), asthma, osteoarthritis, psoriasis, atherosclerosis,
 CC rhinitis, autoimmunity, and organ transplantation, chronic inflammatory
 CC pelvic disease, multiple sclerosis, organ transplantation, myocardial
 CC infarction, cardiovascular disease, stroke, renal failure, oxidative
 CC stress-related neurodegenerative disorders (e.g. amyotrophic lateral
 CC sclerosis, Parkinson's disease and Leigh syndrome), cancer.
 CC cardiomyopathies, ischemic disorders, inflammatory disorders, diabetes
 CC mellitus, fibrotic and mesangial disorders. The proteins may also be
 CC useful for cell growth regulation (e.g. in wound healing), T cell
 CC activation, mitosis control, and as immunosuppressants.
 CC
 XX
 SO Sequence 1239 AA.
 alignment_scores:
 Quality: 96.00 Length: 174
 Ratio: 1.043 Gaps: 5
 Percent Similarity: 52.874 Percent Identity: 21.839
 alignment_block:
 US-09-823-101-4 x AAY55931 ..
 Align seg 1/1 to: AAY55931 from: 1 to: 1239

353 ArgArspheLeuArgLeuGlnGlnLysnLysGluArgSerGluAl 369
 186 ACTACCGGCACACAGCTAGAGAAACACATCTACTAGAACTCAACACAT 235
 369 aLeuArgArgGlnGlnLeuGlnGln.....GlnG 380
 236 ATATACATATAGTACAGAAAGCCACATAGAGCTACTGCA..... 276
 380 InLeuArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 396
 277TAAAGTCAACTGCGCAAGACACGATATAGAGACTGCA 314
 397 LysArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 413
 315 CGAGAGAGAAACGACGACAGAGAGACAAACGACGCGGTGAGC 364
 413 nArgArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 430
 365 GCGATGAGGACAGCGCAGATACCGAAAGAAAGCGGAAAGACAGATG 414
 430 rgGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 446
 415 AAAGTAATCTCAGTTGACGACAGACCAAGCAGACAGCATGTTGATCA 464
 447 Glu.....GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 459
 465 AAGAAACCGCGCGGACAGAC 486
 459 gArgValGlnArgGlnGlnGln 466
 seq_name: /SIDSB/gcgdata/geneseq/geneseq/AA2000.DAT: AAB43016
 seq_documentation_block:
 ID AAB43016 standard; protein; 1165 AA.
 XX
 AC AAB43016:
 XX
 DT 08-FEB-2001 (first entry)
 DT
 XX
 DE Human ORFX ORF2780 polypeptide sequence SRQ ID NO:5560.
 XX
 KW Human: open reading frame; ORFX; detection; cytosolic; hepatotropic;
 KW vulnaray; antiparasitic; antiparkinsonian; nootropic; neuroprotective;
 KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
 KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
 KW antiviral; antibacterial; antifungal; antirheumatic; antihypoid;
 KW antihaemic; gene therapy; cancer; proliferative disorder; hypertension;
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 KW cholesterol ester storage; systemic lupus erythematosus; infection;
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
 KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
 KW thrombosis; contraceptive.
 XX
 OS Homo sapiens.
 XX
 PN WO200058473-A2.
 XX
 PD 05-OCT-2000.
 XX
 PF 31-MAR-2000; 2000WO-US08621.
 XX
 PR 31-MAR-1999; 99US-0127607.
 PR 02-APR-1999; 99US-0127636.
 PR 05-APR-1999; 99US-0127728.
 PR 30-MAR-2000; 2000US-0540763.
 XX
 PA (CURA-) CURAGEN CORP.
 XX
 PI Shinkets RA, Leach M;
 XX

339 CAGAAGAACGACGACAGAGGAGACAAACCAAGAGCCAGCCGUGAGCCCT-

US-09-823-101 from: 1 to: 737

```

alignment_scores:          94.00      Length: 228
Quality:                   0.969      Gaps: 13
Ratio:                     42.544     Gaps: 23.246
Percent Similarity:
Percent Identity:
alignment_block:
US-09-823-101-4 x AAR56494
Align seg 1/1 to: AAR56494 from: 1 to: 737

```

```
67 AAAAAAAAAAAAAAAAACCCGGGGAAACCA..... 102
   ::::: ||| |||: ||| |||
468 GlnLysAsnLysLeuLysGlnProGlyGlySerPheArgAspAspAs 484
103 .....GGGACATAGCGAGTGTCCCGGCTGGAACTGTGTACCCGG 145
   ::::: ||| |||: ||| |||
484 pAspIleAsnAspValAlaSerMetAlaGlyValAsnLeuSerGluIuS 501
146 CTACAAACTC.....CAAAGAGCAAA 168
   ::::: ||| |||: ||| |||
501 eAlaIaArgIleLeuAlaThrAsnSerGluLeuValGlyThrLeuThrArg 517
169 CATTGTCAA.....CGCGCTCCACTACCGCGCACACA 200
   ::::: ||| |||: ||| |||
518 SerCysLysAspGluThrPheLeuLeuGlnAlaProLeuGlnArgAl 534
201 GCTAGAG.....AAACAC.....ATTACTAGAAACTCA 229
   ||| ||| ||| ||| ||| |||
534 eLeuGluIleGlyLysLysHisGlyIleThrGluLeuHisProAspVal 551
230 CAACATATATA...CATAGTACAGAAAGCCACATAGAGTACTCGCA 276
   ::::: ||| |||: ||| |||
551 aISerTYrValSerHisAlaThrGlnGlnArgLeuGlnAsnLeuVal... 566
277 TAAAGTCAACTGCGCAAGACACGCTATCAGAGACTGACGAGAGAA 326
   ::: ||| ||| ||| ||| |||
567 .....GluLysIleSerGluThrAlaGlnGlnLysAs 577
327 C.....GACGACAGAGACAAACCAAGACGACGCGT. 360
   ||| ||| ||| ||| ||| |||
577 nPheSerTYrLysAspAspArgTYrGluGlnAlaSerAspValArgA 594
361 .....GAGCGC 366
594 IaGlnLeuLysPhePheGlnLeuAspGlnIleGlnLysGlnArgLys 610
367 GATGAGCGCAGAGCCGATACCCGAAAGAAAGGAAAGACAGATGAA 416
   ||| ||| ||| ||| ||| |||
611 AspGlnGlnArg.....GluIleLeuMetArgAlaAlaLys 623
417 AGTAATCTCAGTTGACGAGAC..... 438
   ||| ||| ||| ||| ||| |||
623 sSerArgSerArgGlnGlnAspProGluGlnLeuArgLeuLysGlnLysA 640
439 .....CAAAGCGAGAGACGAT... 456
640 IaLysGlnMetGlnGlnGlnLeuAlaGlnMetArgGlnArgAspAla 656
457 .....GTTGATCAAGAAGAAACGACGCGAGACAGA 485
   ::: ||| |||: ||| |||
657 AsnLeuThrAlaLeuAlaIleGlyProArgLysLysArgLysValAs 673
486 CTGT.....GCAGAAGTGGC 501
   ||| ||| ||| ||| ||| |||
673 pCysProGlyProGlySerGlyAlaGlnGlySer 684
```



```

146 CTACAAAGCTC.....CAAAAGAGAAA 168
501 erAlaIaArgIleuAlaThrAsnSerGIuLeuValGIyThrLeuThrArg 517
169 CATTTGCAA.....CGCGCTCAGTACCGCGCACACA 200
518 SerCysLysAspGIuThrPheLeuLeuGlnAlaProLeuGlnArgI 534
201 GCTAGAG.....AAACAC.....ATCTACTAGAAACTCA 229
534 eleuGluIleGIyLysLysHisGIyIleThrGIuLeuHisProAspVal 551
230 CAACATATATA...CATAGTACAGAGAAAGCCACATAGAGCTACTCGCA 276
551 alSerTyValSerHisAlaThrGlnGlnArgLeuGlnAsnLeuVal... 566
277 TAAAGTACAACTGCGCAAGACAGCTATCAGAGACTGACGAGAGAGAAA 326
567 .....GluLysIleSerGIuThrAlaGlnGlnLysAs 577
327 C.....GACGACAGAGAGACAAACCAACGACGACGCGT. 360
577 nPheSerTyLysAspAspArgTyGIuGlnAlaSerAspValArg 594
361 .....GAGCGC 366
594 IaGlnLeuLysPhePheGlnGlnLeuAspGlnIleGlnLysGlnArgLys 610
367 GATGAGCAGAGAGCGCAGATACCCGAAAGAGGGGAAAGACAGATGAA 416
611 AspGIuGlnGlnArg.....GluIleLeuMetArgAlaAla 623
417 AGTAATCTCAGTTGACGAAGAC..... 438
623 sSerArgSerArgGlnGlnAspProGIuGlnLeuArgLeuLysGlnLysA 640
439 .....CAAAAGCAGAGAGACGAT... 456
640 IaLysGIuMetGlnGlnGlnGlnLeuAlaGlnMetArgGlnArgAspAla 656
457 .....GTTGATCAAGAGAAACAGCGCGGAGACAGA 485
657 AsnLeuThrAlaLeuAlaAlaIleGIyProArgLysArgLysValAs 673
486 CTGT.....GCAGAGGTCGC 501
673 pCysProGIyProGIySerGIyAlaGlnGlnLys 684

seq_name: /cgn2_6/prodata/2/1aa/5A_COMB.pep:US-08-646-715-16

seq_documentation block:
; Sequence 16, Application US/08646715
; Patent No. 5637686
; GENERAL INFORMATION:
; APPLICANT: Tjian, Robert
; APPLICANT: Comai, Lucio
; APPLICANT: Dynlacht, Brian D.
; APPLICANT: Hoey, Timothy
; APPLICANT: Ruppert, Siegfried
; APPLICANT: Tanese, Naoko
; APPLICANT: Wang, Edith
; APPLICANT: Weinzierl, Robert O.J.
; TITLE OF INVENTION: NUCLEIC ACIDS ENCODING TAIFS AND METHODS OF USE
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOBAUGH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-4187

```

```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/646,715
; FILING DATE: 09-MAY-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/188,582
; FILING DATE: 28-JAN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Osman, Richard A.
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: A-57650-2/AJT/RAO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 737 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-646-715-16

alignment_scores:
Quality: 94.00 Length: 228
Ratio: 0.969 Gaps: 13
Percent Similarity: 42.544 Percent Identity: 23.246

alignment_block:
US-09-823-101-4 x US-08-646-715-16 ..

Align seg 1/1 to: US-08-646-715-16 from: 1 to: 737

67 AAAAAAAAAAAAAAAAAAAGCGGCGGGAACCA..... 102
468 GlnYsAsnLysLeuLysGIuProGIyGIySerPheArgAspAspAs 484
103 .....GGGACATAGCGAGTGTCCCGCTGGAAGTGTGTACCCGG 145
484 pAspIleAsnAspValAlaSerMetAlaGIyValAsnLeuSerGIuGln 501
146 CTACAAAGCTC.....CAAAAGAGAAA 168
501 erAlaIaArgIleuAlaThrAsnSerGIuLeuValGIyThrLeuThrArg 517
169 CATTTGCAA.....CGCGCTCAGTACCGCGCACACA 200
518 SerCysLysAspGIuThrPheLeuLeuGlnAlaProLeuGlnArgI 534
201 GCTAGAG.....AAACAC.....ATCTACTAGAAACTCA 229
534 eleuGluIleGIyLysLysHisGIyIleThrGIuLeuHisProAspVal 551
230 CAACATATATA...CATAGTACAGAGAAAGCCACATAGAGCTACTCGCA 276
551 alSerTyValSerHisAlaThrGlnGlnArgLeuGlnAsnLeuVal... 566
277 TAAAGTACAACTGCGCAAGACAGCTATCAGAGACTGACGAGAGAGAAA 326
567 .....GluLysIleSerGIuThrAlaGlnGlnLysAs 577
327 C.....GACGACAGAGAGACAAACCAACGACGACGCGT. 360
577 nPheSerTyLysAspAspArgTyGIuGlnAlaSerAspValArg 594
361 .....GAGCGC 366

```

```

594 lAgLInLeuLysPhepHeGInGInLeuAspGlnIleGInLysGlnArgLys 610
367 GATGAGGCGACAGCGAGATACCCGAAGAAGCGGGAAGACAGATGAA 416
||||| ||||| ||||| |||||
611 AspGInGInGInArg.....GlnIleuMetArgAlaAlaLys 623
417 ACTAATCTCAGTTGACGAGAC..... 438
| ||| ::|||
623 sSerArgSerArgGInGInLysAspProGInGInLeuArgLeuLysGInLysA 640
439 .....CAAAACGCAAGAACGAT... 456
||||| ||||| |||||
640 lAlYsGInuMetGInGInGInGInLLeuAlaAlaThrProGInArgGInArgSpAla 656
457 .....GTTGATCAAGAAGAACCGCGGACAGACAGA 485
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
657 AsnLeuThrAlaLeuAlaAlaIleGInProArgLysLysArgLysValAs 673
486 CTGT.....GCAGAGGTGCG 501
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
673 pCysProGInYProGInYSerGInYAlaGInGInYSer 684

seq_name: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:US-08-990-114-3

seq_documentation_block:
: Sequence 3, Application US/08990114
: Patent No. 5932475
: GENERAL INFORMATION:
: APPLICANT: Bandman, Olga
: APPLICANT: Yue, Henry
: APPLICANT: Corley, Neil C.
: APPLICANT: Shah, Purvi
: TITLE OF INVENTION: HUMAN NUCLEOLIN-LIKE PROTEIN
: NUMBER OF SEQUENCES: 3
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Incyte Pharmaceuticals, Inc.
: STREET: 3174 Porter Drive
: CITY: Palo Alto
: STATE: CA
: COUNTRY: USA
: ZIP: 94304
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: OPERATING SYSTEM: DOS
: SOFTWARE: FastSeq for Windows Version 2.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/990.114
: FILING DATE: Herewith
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER:
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Billings, Lucy J.
: REGISTRATION NUMBER: 36,749
: REFERENCE/DOCKET NUMBER: PF-0451 US
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 650-855-0555
: TELEFAX: 650-845-4166
: TELEX:
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 714 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: IMMEDIATE SOURCE:
: LIBRARY: GenBank
: CLONE: 128842
: US-08-990-114-3

```

```

alignment_scores:
  Quality: 91.50      Length: 170
  Ratio: 0.943      Gaps: 7
  Percent Similarity: 57.059      Percent Identity: 24.118

```

alignment_block:

US-09-823-101-4 x US-08-990-114-3 ..

Align seg 1/1 to: US-08-990-114-3 from: 1 to: 714

```

38 AAAAAAGTAAATCTATCAAGAGAAAAA.....AA 69
||||| ||||| ||||| |||||
61 LysLysValValValSerGInThrLysValAlaValProThrProAl 77
70 AAAAAAAAAAAAAAAAAACCGGGGGAACCAAGGAGCATAGCGAGTGT 119
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
77 alYsLysAlaAlaValThrProGInYsLysAlaAlaIleThrProAlL 94
120 CCGGGGCTGGAAACTGTGTACCGGCTACAAACTCCA...AAGAGAGA 166
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
94 yAlYsAla.ValThrProAlLysAlaValAlaIleThrProGInYsLysG 110
167 AACATTGTCAAGCGCGCTCCACTACGCGCACACAGCTAGAGAACATC 216
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
110 yAlaThrGInAlaLysAlaLeuValAlaThrProGInYsLysGInYAla 127
217 TACTGAAGAACTCAACATATATATACATAGAGAGAAACGCACATAGA 266
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
127 al.....ThrProAlLys 131
267 GCTACTCGCATTAAGTACACTGCGCAAGACACGTATACAGACTGACG 316
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
132 GInYAlaLysAsnGInYLysAsn.AlalYsLysGInLysPserAspGIn 148
317 AGAGAAAGAAAGACGACGACGAGAGCAACCAAGCAAGCGGAGTGAGCC 366
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
148 lAspAspAspAspAspGInLysPserAspGInLysPserGInLysPser 164
367 GATGAGGCGACAGCGAGATACCCGAAGAAGCGGGAAGACAGATGAA 416
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
165 GInGInAspGInPheGInuProValValLysGInLys...GInGInLys 180
417 AGTA.....ATCTCAGTTGACGAAGACCAAAAGCGAGAG 451
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
180 sValAlaAlaAlaProAlaSerGInLysPserGInLysPserGInLysP 197
452 ACGATGTTGATCAAGAAACGCGGACAGACAGACTGTGCAGAGGTGCG 501
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
197 lGInGInGInGInGInLysPserGInGInGInLysPserGInGInLys 213
502 AGGCG 507
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
214 GInAla 215

seq_name: /cgn2_6/ptodata/2/1aa/6B_COMB.pep:US-09-370-368-9

seq_documentation_block:
: Sequence 9, Application US/09370368
: Patent No. 6258932
: GENERAL INFORMATION:
: APPLICANT: Anders Valine
: TITLE OF INVENTION: PEPTIDES THAT BLOCK VIRAL INFECTIVITY
: FILE REFERENCE: TRIPEP.003A
: CURRENT APPLICATION NUMBER: US/09/370.368
: CURRENT FILING DATE: 1999-08-09
: NUMBER OF SEQ ID NOS: 10
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 9
: LENGTH: 538
: TYPE: PRT
: ORGANISM: Moloney Murine Leukemia Virus
: US-09-370-368-9

```

alignment_scores: Quality: 91.00 Length: 163
 Ratio: 1.071 Gaps: 9
Percent Similarity: 52.147 Percent Identity: 26.380

alignment_block:

US-09-823-101-4 x US-09-370-368-9 ..

Align seg 1/1 to: US-09-370-368-9 from: 1 to: 538

```
84  AAACCCGGGGGGAACACGAGGAGCATACCGAGTCCCGG.....CT 127
   |||:||||| |||:||||| |||:||||| |||:||||| |||:
353  LysValLysGlyIleThrGlnGlyProAsnGlnSerProSerAlaPheLe 369
128  GGAACCTGTGTACCCGGCTACAA.....A 153
   |||: |||: |||: |||: |||: |||: |||: |||: |||:
369  uGluArgLeuLysGlnAlaTyrArgTyrThrProTyrAspProGluA 386
154  CTCCAAAAGAGAAACATTTTCACCGCGCTCCACCTACCGCACACAGCT 203
   |||: |||: |||: |||: |||: |||: |||: |||: |||:
386  sPProGlyGlnGluThrAsnValSerMetSerPheIleTyrGlnSerAla 402
204  .....AGAGAACACATCTACTA..... 221
   |||: |||: |||: |||: |||: |||: |||: |||: |||:
403  ProAspIleGlyArgGlyLeuGlnArgLeuLysPLeuArgAsnLysTh 419
222  .....GAACCTCACACATATATACATAGTACAG 252
   |||: |||: |||: |||: |||: |||: |||: |||: |||:
419  rLeuGlyAspLeuValArgGlnAlaGluArgIlePheAsn...LysArg 435
253  AAACGCCACATAGAGCTACTGCATTAAGTCAACTGCGCAAGACAGACT 302
   |||: |||: |||: |||: |||: |||: |||: |||: |||:
435  LuthrProGluGlu.....ArgGlnGluArg 443
303  ATCAGAGACTACGAGAGAAACGACGACAGAGAGACAAACACAGCA 352
   |||: |||: |||: |||: |||: |||: |||: |||: |||:
444  IleArgArgGlnArgGlnGluGlnGlnGlnArgArgThrGlnAspGln 460
353  GCCAGCGTGAAGCGGATGAGCGGACAGATACCCGAAGAAACGGGG 402
   |||: |||: |||: |||: |||: |||: |||: |||: |||:
460  u.GlnLysGlnLys...GlnArgAspArgArgGlnHisArgGlnMetSer 475
403  AAAAGACAGATGAAGTAATCTCACTTGACGAAAGCCAAAGCGAGAGA 452
   |||: |||: |||: |||: |||: |||: |||: |||: |||:
476  ArgLeuLeuAlaThrValValSer.....GlyGlnArgGlnAspArg 489
453  CGATGTTGATCAAGAAACCGACCGCGAGACGAGACTGT 489
   |||: |||: |||: |||: |||: |||: |||: |||: |||:
489  gGlnGlnGlnGlnGlnArgArgArgSerGlnLeuAspCys 501
```

seq_name: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:US-09-393-569-2

seq_documentation_block:

```
; Sequence 2, Application US/09393569
; Patent No. 6277979
; GENERAL INFORMATION:
; APPLICANT: BINGHAM, SHARON
; APPLICANT: CASE, PATRICK
; APPLICANT: LAWSON, SALLY NEALE
; APPLICANT: NEWTON, RICHARD ANTHONY
; APPLICANT: PIERCEY, VALERIE
; APPLICANT: RAUSCH, OLIVER LARS
; APPLICANT: RAVALL, PRAVIN
; APPLICANT: REITH, ALASTAIR DAVID
; APPLICANT: SANGER, GARETH JOHN
; TITLE OF INVENTION: NEW USE
; FILE REFERENCE: P32261
; CURRENT APPLICATION NUMBER: US/09/393,569
; CURRENT FILING DATE: 1999-09-10
; EARLIER APPLICATION NUMBER: GB 9907261.3
; EARLIER FILING DATE: 1999-03-29
```

```
; EARLIER APPLICATION NUMBER: GB 9819779.1
; EARLIER FILING DATE: 1998-09-10
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1360
; TYPE: PRT
; ORGANISM: HOMO SAPIENS
US-09-393-569-2
```

alignment_scores: Quality: 87.00 Length: 168
 Ratio: 1.048 Gaps: 6
Percent Similarity: 49.405 Percent Identity: 24.405

alignment_block:

US-09-823-101-4 x US-09-393-569-2 ..

Align seg 1/1 to: US-09-393-569-2 from: 1 to: 1360

```
43  CGTATTTCTATCAAGAGGAAAAA..... 84
   |||: |||: |||: |||: |||: |||: |||: |||: |||:
303  LysAspHisIleAspArgThrLysLysArgGlnGlnLysAspGlnThr 319
85  .....AAACCCGGGG 94
   |||: |||: |||: |||: |||: |||: |||: |||: |||:
319  rGlnTyrGlnTyrSerGlnSerGlnGlnGlnGlnGlnGlnAsnSerG 336
95  GGAACACGAGGAGCATAGCGAGTGTCCGGGCTGAAA...CTGTGGTAC 141
   |||: |||: |||: |||: |||: |||: |||: |||: |||:
336  LysLysProSerSerIleLeuAsnLeuProGlyGlnSerThrLeuArgArg 352
142  CCGGCTACAAACCTCCAAAGAGAAACATTTGTCAACGCGCTCCACTACC 191
   |||: |||: |||: |||: |||: |||: |||: |||: |||:
353  AspPheLeuArgLeuGlnLeuAlaAsnLysGlnArgSerGlnAlaLeuAr 369
192  GCGCACACAGCTAGAGAAACATCTACTAGAACTCACAACATATATAC 241
   |||: |||: |||: |||: |||: |||: |||: |||: |||:
369  gArgGlnGlnLeuGlnGlnGln.....GlnA 378
242  ATAGTACAGAAAGCGCACATAGAGCTACTCGCATTAAGTACAACTGGC 291
   |||: |||: |||: |||: |||: |||: |||: |||: |||:
378  rGlnLysGlnGlnGlnLysArgGlnLeuLeuAla.....GlnArg 391
292  CAAAGACAGTATGAGAGCTAGAGAGAGAAAGAACGACGACAGAGAGA 341
   |||: |||: |||: |||: |||: |||: |||: |||: |||:
392  GlnLysArgIleGlnGlnLysGln..... 400
342  CAAACCAACGACCGGCTGAGCGCATGAGGAGAGCGCAGATTACCGCA 391
   |||: |||: |||: |||: |||: |||: |||: |||: |||:
401  .....GlnArgArgArgLeuGlnGlnGlnGlnArgArgGln 413
392  AAGAACGCGGAAAGACAGATGAAGTAATCTCAGTTGACGAGAACCA 441
   |||: |||: |||: |||: |||: |||: |||: |||: |||:
413  ysgIleuArgLysGlnGln.....GluArgGlnGln 423
442  AAGCGAGAGACGATGTTGATCAAGAAAGAACGAGCGGACAGACTGCG 491
   |||: |||: |||: |||: |||: |||: |||: |||: |||:
424  ArgArgHisIstyrGlnGlnGlnMetArgArgGlnGlnArgArgArgAl 440
492  AGAA 495
   |||: |||: |||: |||: |||: |||: |||: |||: |||:
440  aglu 441
```

seq_name: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:US-08-903-801-1

seq_documentation_block:

```
; Sequence 1, Application US/08903801
; Patent No. 5932712
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer
; APPLICANT: Corley, Neil C.
```



```

APPLICANT: Shah, Purvi
TITLE OF INVENTION: NEW ANEXIN BINDING PROTEIN
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/903,801
FILING DATE: Herewith
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0354 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 290 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: PROSNON01
CLONE: 2272281
US-08-903-801-1

alignment_scores:
Quality: 86.50 Length: 143
Ratio: 1.095 Gaps: 5
Percent Similarity: 55.245 Percent Identity: 25.175

alignment_block:
US-09-823-101-4 x US-08-903-801-1 ..

Align seg 1/1 to: US-08-903-801-1 from: 1 to: 290

54 CAAGAGGAAAAAAAAAAAAAAAAAACC CGGGGGAACGAG 103
   :::::::::::::::::::: :::::::::::
149 LysLysGluArgLysLysArgAspGluLysAlaLysLeuAr 165
   :::::::::::::::::::: :::::::::::
104 GGGACATAGCGGTGCCGGGCTGGAACGTGTGATCCCGCTACA 153
   :::::::::::::::::::: :::::::::::
165 GlySerLeuLysGluLysGluLysLeuGluThr.....GlyLysAs 179
   :::::::::::::::::::: :::::::::::
154 CTCCAAGAGGAACATTGTCACGCGCTCCACTACCGCGCACAGCT 203
   :::::::::::::::::::: :::::::::::
179 spGlnSerLysGlnLysGlnSerGlnArgLysPheGluLysLysVal 195
   :::::::::::::::::::: :::::::::::
204 AGAGAAACATCTACTAGAAACTCACAACATATATACATACTACAGAGA 253
   :::::::::::::::::::: :::::::::::
196 LysSerLysValThrValAspThrGlyValIleProAla.SerGluGluL 212
   :::::::::::::::::::: :::::::::::
254 AACGCCCATAGAGCTACTCGCATTAAGTACAACTCGCGCAAGACAGCTA 303
   :::::::::::::::::::: :::::::::::
212 Ys.....AlaGluThrProThrAlaIleGluLysPsp 222
   :::::::::::::::::::: :::::::::::
304 TCAGAGACTGACGAGAGAAAGACGACGACGAGAGAGCAACCAACGAG 353

```

```

223 AsnGluGlyAspLysLysLys.....AspLysLysLys 235
   :::::::::::::::::::: :::::::::::
354 CCAGCGTGAGCGCGATGAGCGAGCGAGCATACCCGAAGAAACGGGGA 403
   :::::::::::::::::::: :::::::::::
235 LysSerLysGluLysGluLysGluLysLysLysLysLysLysLysLys 252
   :::::::::::::::::::: :::::::::::
404 AAAGACAGATGAAGTATC.....TCAGTT 429
   :::::::::::::::::::: :::::::::::
252 ysaLatThrValLysAlaMetGlnGluAlaLeuAlaLysLeuLysGlu 268
   :::::::::::::::::::: :::::::::::
430 GACGAGACCAAAACGAGAGACAGCAT 456
   :::::::::::::::::::: :::::::::::
269 GluGluArgGlnLysArgLysGluLys 277
   :::::::::::::::::::: :::::::::::

seq_name: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:US-09-295-055-1

seq_documentation_block:
Sequence 1, Application US/09295055
Patent No. 6232440
GENERAL INFORMATION:
APPLICANT: Hillman, Jennifer
APPLICANT: Corley, Neil C.
TITLE OF INVENTION: NEW ANEXIN BINDING PROTEIN
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/295,055
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/903,801
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0354 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 290 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: PROSNON01
CLONE: 2272281
US-09-295-055-1

alignment_scores:
Quality: 86.50 Length: 143
Ratio: 1.095 Gaps: 5
Percent Similarity: 55.245 Percent Identity: 25.175

alignment_block:
US-09-823-101-4 x US-09-295-055-1 ..

```

Align seg 1/1 to: US-09-295-055-1 from: 1 to: 290

```

54 CAAGAGGAAAAAAAAAAAAAAAAAACCGGGGGAACACG 103
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:
149 Lysylserglylvalarglylsaraspelulysalalysleu 165
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:
104 GGGACATAGCAGTGTCCCGGCTGGAACCTGTGTACCCGGCTACAAA 153
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:
165 glyleuleylsglulsgluleuylthr.....Glylylsysa 179
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:
154 CTCCAAAAGAGAAACATTGTCAACGGCTCCACTACCGGCACACACT 203
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:
179 spglnserlysglnlysglnserlgnarplysnehluglulthrval 195
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:
204 AGACAACACATCTACTAGAACTCACACATATATATAGTACAGAGA 253
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:
196 Lyserylvalthrvalaspthrlyvalleproala.Serlulglul 212
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:
254 AACGCCACATAGAGCTACTCGCATTAAGTCAACTGCCAAGACAGTA 303
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:
212 ys.....Alagluthrprothrlealaglualaspasp 222
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:
304 TCAGAGACTACGAGAGAGAAACGACAGAGAGAGCAAAACACAGAG 353
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:
223 Asnlgulglulysplyslylslys.....Asplyslylslys 235
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:
354 CCACGCTGAGCGGATGAGCGAGCGCAGATACCCGAAAGAAACGGGGA 403
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:
235 slvslyglulglulsglulglulsglulsglulsglulsglulsgl 252
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:
404 AAACAGACATGAAAGTAATC.....TCAGTT 429
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:
252 ysalaThrvallysalametcglulglulaleuallalysleulsglul 268
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:
430 GACGAGAACCAAAAGCGAGAGAGAT 456
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:
269 glulglularglulglulysargylulglul 277
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:

```

seq_name: /cgn2_6/ptodata/2/1aa/6B_COMB.pep:US-08-942-008-2

```

seq_documentation_block:
; Sequence 2, Application US/08942008
; Patent No. 6133419
; GENERAL INFORMATION:
; APPLICANT: Braselmann, Sylvia
; TITLE OF INVENTION: Nucleotide Sequences that Encode
; TITLE OF INVENTION: Phosphatidylinositol-3' Kinase Associated Proteins and
; TITLE OF INVENTION: Uses thereof
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ONYX Pharmaceuticals, Inc.
; STREET: 3031 Research Drive
; CITY: Richmond
; STATE: CA
; COUNTRY: USA
; ZIP: 94806
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/942,008
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Giotta, Gregory
; REGISTRATION NUMBER: 32,028
; REFERENCE/DOCKET NUMBER: ONYX1027
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 262-8710
; TELEFAX: (510) 222-9758
; INFORMATION FOR SEQ ID NO: 2:

```

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 589 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-942-008-2

```

```

alignment_scores:
  Quality: 86.00      Length: 100
  Ratio: 1.509      Gaps: 2
  Percent Similarity: 57.000      Percent Identity: 23.000

```

alignment_block:

US-09-823-101-4 x US-08-942-008-2 ..

Align seg 1/1 to: US-08-942-008-2 from: 1 to: 589

```

205 GAGAAACACATCTACTAGAACTCACACATATATACATAGTACAGAA 254
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:
11 AsplyshisleuYr.....GlulglulThrvalglulysproleu 24
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:
255 ACCGCCACATA.....GAGCTACTCGCATTAAGTACACTG 289
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:
24 sleuvalleuylvalgllylsnglulvalThrglulenseThrclys 41
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:
290 CGCAAAACACGCTATCTGAGACACTGACGAGAGAAAGAAACGACACAGGA 339
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:
41 erSerlyshisaspserSerleupheglulaspIysasnaspHisasp 57
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:
340 GACAAACCAAGAGCCAGCGTGAAGCGGATGAGGACGAGAGCCAGANTACC 389
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:
58 Hislysasparglysarplysarglylslysglylulysglnlepr 74
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:
390 GAAAGAAACGGGGGAAAGACAGATGAAGTAATCTCAGTTGACAGAGACC 439
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:
74 oglyglulglulysgllyarglysarparargvallysgluaspIysly 91
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:
440 AAAAGCGAGAAAGCAGATGTGATCAAAAGAAACGCGCGGACGAGACTGT 489
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:
91 ysarigaspargaspargvalglulasmglulalaglulysaspIeuglly 107
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:
seq_documentation_block:
; Sequence 24, Application US/07955905A
; Patent No. 5770433
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: RECOMBINANT 47 AND 31 KD COCOA PROTEINS AND
; TITLE OF INVENTION: PRECURSOR
; NUMBER OF SEQUENCES: 28
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/955,905A
; FILING DATE: 21-JAN-1993
; CLASSIFICATION: 435
; INFORMATION FOR SEQ ID NO: 24:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 605 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: Protein
; LOCATION: 1..605
; OTHER INFORMATION: /note= "Vicilin from G. max"

```

US-07-955-905A-24

Alignment_scores:
Quality: 85.50 Length: 152
Ratio: 1.096 Gaps: 8
Percent Similarity: 51.316 Percent Identity: 25.658

Alignment_block:

US-09-823-101-4 x US-07-955-905A-24 ..

Align seg 1/1 to: US-07-955-905A-24 from: 1 to: 605

```
66 AAAAAAAAAAAAAAAAAACCGGGGAA.....ACGAGGG 106
111 : : : : : : : : : : : : : : : : : : : : :
62 LysValGluLysGluGluGluGluGluGluLeuProArgProArgPr 78
107 ACATAGCGAGTGTCCCGGCTGGAACCTGGTACCCGGCTACAAACTC 156
78 oArGProGlnHisProGluArgGluProGlnProGluGluLysGluG 95
157 CAAAGAGGAACATTGTCAAGC...GCTCCACTACCGGACACAGCT 203
95 LuAspGluAsp.GluGlnProArgProIleProPheProArgProGlnPr 111
204 AGAAGAACACATCTACTAGAACTCACAACATATATACATAGTACAGA 253
111 oArgGln.....GluG 115
254 AACGCCACATAGAGCTACTCGCATAAAGTACAACCTCCGAAGACAGCTA 303
115 LuGlnHis.....GluGlnGluGluGln 122
304 TCAGAGACTGACGAGAGAAAGACGACGACGAGAGAACAAACGAG 353
123 GlnGluTTP..ProArgLysGluGluLysArgGlyGluLysGlySerG 138
354 CCAGCGTGAGCGCGATGAG.....GCAGACGCGACATACCCGA 391
138 uGluGluAspGluAspGluAspGluGluGlnAspGluArgGlnPhePro 155
392 AAGAAACGGGAAAGACAGATGAAGTATCTCACTTACGACAGACCAA 441
155 heProArgProProHisGln.....LysGluGluArg 165
442 AAGCGGAGAGAGATGTGTCAAGAAACGACGCGAGACAGACGCTGC 491
166 AsnGluGluGluAspGluAspGluGluGlnArgGluSerGluLysE 182
492 AGAA 495
182 rGlu 183
```

seq_name: /cgn2_6/ptodata/2/iaa/5B-COMB.pep:US-08-933-750C-47

seq_documentation_block:
: Sequence 47, Application US/08933750C
: Patent No. 5932442
: GENERAL INFORMATION:
: APPLICANT: Lal, Preeti
: APPLICANT: Hillman, Jennifer L.
: APPLICANT: Bandman, Olga
: APPLICANT: Shah, Purvi
: APPLICANT: Au-Young, Janice
: APPLICANT: Yue, Henry
: APPLICANT: Guegler, Karl J.
: APPLICANT: Corley, Neil C.
: TITLE OF INVENTION: HUMAN REGULATORY MOLECULES
: NUMBER OF SEQUENCES: 98
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Incyte Pharmaceuticals, Inc.
: STREET: 3174 Porter Drive
: CITY: Palo Alto

```
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/933,750C
FILING DATE: September 23, 1997
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0356 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX:
INFORMATION FOR SEQ. ID NO: 47:
SEQUENCE CHARACTERISTICS:
LENGTH: 432 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: TLYMN0704
CLONE: 2926777
US-08-933-750C-47
```

Alignment_scores:
Quality: 82.00 Length: 159
Ratio: 0.943 Gaps: 6
Percent Similarity: 54.717 Percent Identity: 23.270

Alignment_block:
US-09-823-101-4 x US-08-933-750C-47 ..

Align seg 1/1 to: US-08-933-750C-47 from: 1 to: 432

```
15 CAGAGTGTATATCATCGAGGAAACGTAATCTATCAAGAGGAAA 64
243 ArgAspGluArgLeuLysGluLysGlnGluArgGluGluArgGluL 259
65 AAAAAAAAAAAAAAAAAACCGGGGGAACACAGGCGACATAGCG 114
1: : : : : : : : : : : : : : : : : : : : :
259 sGluArgGluArgGluArgGluGluGluArgGluArgGluArgGluG 276
115 AGTGTCGCGGCTGGAACCTGTGTACCCGCTACAAACTCCAAAGAG 164
276 Lu.....GluGluArgGluLysGluArgGluArgGluArgGlu 288
165 GAACACTTGTCAAGCGCTCCACTACCGCGCACACGCTAGAGAAACACA 214
289 ArgArgLysArgSerArgSerArgSerArgHisSerArgThSerAs 305
215 TCTACTAGAACTCACACATATATACATAGTACAGAGAACGCCACATA 264
305 parArgArgCysSer.....ArgSerArgAspHisL 315
265 GAGCTACTCGCATTAAGTACAACCTGCGCAAGACACGATTCAGAGACTGA 314
: : : : : : : : : : : : : : : : : : : : :
315 yArgSerArgSerArgGluArg...ArgArgThArgSerArgSparG 330
315 CGAGAGA...AGAAAGCAGCAGAGAGAGAGCAAAACACAGACGACGCG 361
331 Arg.ArgSerArgSerHisAspArgSerGluArgGlyHisArgSerArgS 347
```



```

; FILING DATE: HEREMITH
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0465 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-855-0555
; TELEFAX: 650-845-4166
;
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 688 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: HEARNOT01
; CLONE: 307624
;
; US-09-016-000-2

```

```

alignment_scores:
  Quality: 81.00      Length: 164
  Ratio: 1.025        Gaps: 7
  Percent Similarity: 48.171  Percent Identity: 25.000

```

```
alignment_block:
US-09-823-101-4 x US-09-016-000-2 ..
```

```
Align seg 1/1 to: US-09-016-000-2 from: 1 to: 688
```

```

19 ATGTGATATCATCAGGAAAAAAGCTATTCATATCAGAGGAAAAA 68
   |||||||:|||||:|||||:|||||:|||||:|||||:|||||:
222 MetCysValAspAspAlaTyrValArgArgMetAlaAlaGluAlaThrGlu 238
69 AAAAAAAGGAAAAAAGGAAAAAAGGAAAAAAGGAAAAAAGGAAAAA 118
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
238 utrglnlslsalaaglyalaproprosergylseralalavaserthra 255
119 TCCCGGGCTGAAACTGTGTACCGGCTACAAAGCTCCAAAGAGAGAA 168
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:
255 lalproglnglnlsls.....Proileglylslsleaserlslsaslsls 268
169 CATGTGCAACGGCTCCACTACCGGCGACACAGACTAGAGAAACACATCTA 218
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:
269 .....LysLysLysLeulLysLys..... 274
219 CTGAAACTCACACATATATACATAGTACAGAGAAAGCCACATAGAGC 268
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
275 ...Lysglnlslsthrglnalaglualeuenglnlslsargleuglnlsl 290
269 TACTCGCATTAAGTACACTCGCAAGAACACATATCAGAGACTGAC... 315
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:
290 leglucluleuengluarargluAlaIagluasglleilegluIuasnile 306
316 .....GAGAGAAAGAAAGCGACAGAGAGAC..... 342
307 ThSerAlaAlaProSerAsnAspGlnAspGlyGluTyrCysProGluVa 323
343 .....AAACCAAGCAGCGAGCGGCGGAGGAGGAGGAGGAGAGAT 385
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
323 lylslslslsthrgllyleuglnlslualalagluAlaIagluThr... 338
386 ACCGAAAGAAAGCGGAAAGAGAGAGATGAAGTAATCTCAGTTGAGAA 435
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
339 ..Alalysaspasnlglylualagluaspqlnglnlslslslslslslsl 354
436 GAG.....CAAAAGCAGAAAGAGATGTTGATCAA 465
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```
355 GlulysGluAsnIleGluLysAspGluAspAlaIaspGln 368
```

```
seq_name: /cgn2_6/ptodata/2/1aa/6B_COMB.pep:US-09-245-041-11
```

```
seq_documentation_block:
```

```
; Sequence 11, Application US/09245041
```

```
; Patent No. 6274339
```

```
; GENERAL INFORMATION:
```

```
; APPLICANT: Moore, K.
```

```
; APPLICANT: Nagle, D.
```

```
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE DIAGNOSIS AND TREATMENT
```

```
; TITLE OF INVENTION: OF BODY WEIGHT DISORDERS INCLUDING OBESITY
```

```
; FILE REFERENCE: 7853-136
```

```
; CURRENT APPLICATION NUMBER: US/09/245,041
```

```
; CURRENT FILING DATE: 1999-02-05
```

```
; EARLIER APPLICATION NUMBER: 60/093,630
```

```
; EARLIER FILING DATE: 1998-07-21
```

```
; EARLIER APPLICATION NUMBER: 60/104,978
```

```
; EARLIER FILING DATE: 1998-10-20
```

```
; NUMBER OF SEQ ID NOS: 131
```

```
; SOFTWARE: FastSeq for Windows Version 3.0
```

```
; SEQ ID NO 11
```

```
; LENGTH: 351
```

```
; TYPE: PRT
```

```
; ORGANISM: Mus musculus
```

```
; FEATURE:
```

```
; NAME/KEY: SITE
```

```
; LOCATION: all Xaa positions
```

```
; OTHER INFORMATION: Xaa=unknown amino acid
```

```
; US-09-245-041-11
```

```

alignment_scores:
  Quality: 80.50      Length: 44
  Ratio: 2.516        Gaps: 2
  Percent Similarity: 72.727  Percent Identity: 47.727

```

```
alignment_block:
US-09-823-101-4 x US-09-245-041-11 ..
```

```
Align seg 1/1 to: US-09-245-041-11 from: 1 to: 351
```

```

37 AAAAAAGTATTCATCAAGAGGAAAAAAGGAAAAAAGGAAAAA 86
   |||||||:|||||:|||||:|||||:|||||:|||||:|||||:
29 LysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 45
87 ACCCGGGGGAACCAAGGAGACATAGGAGTGTCCGCGGCTGGAACGTG 136
   ||||||:|||||:|||||:|||||:|||||:|||||:|||||:
45 slslslslslslslslslslslslslslslslslslslslslslslsl 59
137 GGTACCGGCTACAAACTCCAAAGAGAGAA 168
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:
60 ..PheAlaIglArgArgLeuSerProArgLys 69
seq_name: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:US-08-557-309B-43
seq_documentation_block:
; Sequence 43, Application US/08557309B
; Patent No. 5916572
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Houghton, Raymond L.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DETECTION AND
; NUMBER OF SEQUENCES: 69
; CORRESPONDENCE ADDRESS:
; ADDRESSSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092

```

```
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/557,309B
FILING DATE: 14-NOV-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.4222
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 43:
SEQUENCE CHARACTERISTICS:
LENGTH: 186 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-08-557-309B-43
```

```
alignment_scores:
Quality: 80.00      Length: 72
Ratio: 1.600        Gaps: 2
Percent Similarity: 69.444      Percent Identity: 26.389
```

alignment_block:

US-09-823-101-4 x US-08-557-309B-43 ..

Align seg 1/1 to: US-08-557-309B-43 from: 1 to: 186

```
289 GCCCAAGACACGTATCAGAGCTGACGAGAGAAGAAACGACGACGAGG 338
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
71 AlArArGrGlUaLaGlUGlUaRgLaLaRgGlUaLaGlUGlUaRgLa 87
339 AGACAACCAACGACGACGCGCGATGAGCGGCGATGAGCGGCGATACC 388
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
87 aGlnArGrGlUaLaGlUGlUaRgLaLaRgGrGlUaLaGlULySaRgLa 104
389 CGAAAGAAACGGGGAAGACACAGATGAAGTAATCTCAGTTGACGAAG 438
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
104 rGrArGrGlUaLaLySGlUaRgLaLaTrpLn.....GlUaLaGlU 116
439 CAAAGCGAGAGACGATGTTGATCAAGAAACGACGCGGACGACAGCTG 488
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
117 GlUaRgLaLaGlInArGrGlUaLaGlUGlUaRgLaLaRgArGrGlUa 132
489 TGCAGAAGGTGCGAGG 504
||| |||||
133 ....GlUaRgLaLaRg 136
```

seq_name: /cgn2_6/ptodata/2/1aa/6A_COMB.pep:US-08-834-306-43

seq_documentation_block:

Sequence 43 Application US/08834306

Patent No. 6054135

GENERAL INFORMATION:

APPLICANT: Reed, Steven G.

APPLICANT: Skelky, Yasir A.W.

APPLICANT: Lodes, Michael J.

APPLICANT: Houghton, Raymond L.

TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DETECTION AND

NUMBER OF SEQUENCES: 65

CORRESPONDENCE ADDRESS:

ADDRESS: SEED AND BERRY LLP

STREET: 6300 Columbia Center, 701 Fifth Avenue

CITY: Seattle

STATE: Washington

COUNTRY: USA

```
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/834,306
FILING DATE: 15-APR-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.422C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 43:
SEQUENCE CHARACTERISTICS:
LENGTH: 186 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-08-834-306-43
```

```
alignment_scores:
Quality: 80.00      Length: 72
Ratio: 1.600        Gaps: 2
Percent Similarity: 69.444      Percent Identity: 26.389
```

alignment_block:

US-09-823-101-4 x US-08-834-306-43 ..

Align seg 1/1 to: US-08-834-306-43 from: 1 to: 186

```
289 GCCCAAGACACGTATCAGAGCTGACGAGAGAAGAAACGACGACGAGG 338
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
71 AlArArGrGlUaLaGlUGlUaRgLaLaRgGlUaLaGlUGlUaRgLa 87
339 AGACAACCAACGACGACGCGCGATGAGCGGCGATGAGCGGCGATACC 388
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
87 aGlnArGrGlUaLaGlUGlUaRgLaLaRgGrGlUaLaGlULySaRgLa 104
389 CGAAAGAAACGGGGAAGACACAGATGAAGTAATCTCAGTTGACGAAG 438
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
104 rGrArGrGlUaLaLySGlUaRgLaLaTrpLn.....GlUaLaGlU 116
439 CAAAGCGAGAGACGATGTTGATCAAGAAACGACGCGGACGACAGCTG 488
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
117 GlUaRgLaLaGlInArGrGlUaLaGlUGlUaRgLaLaRgArGrGlUa 132
489 TGCAGAAGGTGCGAGG 504
||| |||||
133 ....GlUaRgLaLaRg 136
```


THIS PAGE BLANK (USPTO)


```

409 CAGTGAAGAAGTAAATCTCAGTTGACAGAACCCAAAGCAGACGATG 458
      ||| ||||| ::: ||| |||:
501 VAIgLuLySGLIhISaRTYtFASpGLuAgRIgHISaRTYtYVaLaSpmE 517
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
459 TGAATCAAAAGAAACCCAGGCCCGACACAGAC 486
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
517 tGluSerGluASnArgaSnARgSerGlu 526

seq_name: p1r2:727574

```

seq_documentation_block:
hypothetical protein ZC443.4 - *Caenorhabditis elegans*
C:Species: *Caenorhabditis elegans*
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T27574
R:Baynes, C.
submitted to the EMBL Data Library, June 1996
A:Reference number: Z20389
A:Accession: T27574
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-389 <full>
A:Cross-references: EMBL:T275553; PTDN:CAA99946.1; GSPDB:GN000023; CESP:ZC443.4
A:Experimental source: Clone ZC443
C:Genetics:
A:Gene: CESP:ZC443.4
A:Map position: 5
A:Introns: 114/3; 151/3

[illegible]

```

|||||.....||| |||: .....|||
272 luaspqluglucglutluaspaspspspspspspsglugluglu 286
484 GACTGTGCAGAA 495
:: :|||
289 Gluasnvalglu 292

seq_name: pir2:T03045
```

seq_documentation_block:
hypothetical protein 009R - Chilo iridescent virus
C:Species: Chilo iridescent virus
C:Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 04-Mar-2000
C:Accession: F03045
R:Bahr, U.; Tidona, C.A.; Darai, G.
Virus Genes 15, 235-245, 1997
A:Title: The DNA sequence of Chilo iridescent virus between the genome coordinates 0
A:Reference number: Z14834; MUID:96141693
A:Accession: F03045
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-456 <BAH>
A:Cross-references: EMBL:AF003534; NID:g2738385; PIDN:AAB94419.1; PID:g2738392
C:Superfamily: Chilo iridescent virus hypothetical protein 009R

```

alignment_block:
US-09-823-101-4 x T03045  ..

Align seg 1/1  to: T03045  from: 1  to: 456

51  TATCAAGAGGAAAAAAAAAAAAAAAAAAAAAAAAACCGGGGGGAAC 100
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
30  TyrLeuGluLeuLeuGluAsnLysAsnLysValLysLeu..... 42

101 CAGGGGACATATGAGCAGTGTCCCGGGCGGAAACTGTG..... GTACCCG 144
    43  .....AsnCyS11eGlyLysGluPheValProthThrProP 55
    145 GCTACAAACTCTCAAAAGAGGAAACATTTGTA.....ACG 179
        |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
55  rOlenseThrGlyArgGluThr1LevalSerLysasprOHisValVal 71
    180 CGGTCCACTACCGCGCACACAGCTAGAGAAACATCTACTAGAAACTCA 229
        |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
72  GlnsrSer11eSerAsnThrPro1LeargThc1u11eLysasphThr 88
    230 CAACATATATCATATGATACAGAGAAAGCCACATAGAGCTACTCGCATTA 279
        |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
88  oArg.....TyrGluGluThrPro1LeLysArgThr11eThry 101
    280 AGTACAACTGC.....GCA 293
        |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
101  alThrasValLysThrValLysSerSerSer11eSerGlymetasngly 117
    294 AAGACACGATATCAGAGACTGACGAGAGAGAAACGACGACAGAGAGAC 343
        |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
118  ArgasnArgLeuTyrAspAspAspLeuPheaspAspArgTyrLys 134
    344 AACCAAGAGCCACCGT.....GAGCGGCATGAGCGACAGCCG 381
        |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
134  erProthThrArgLysPheGlnGlyLysArgAspGluAsp11eArg 150
    382 AGATACCGGAGAGAAAG.....GGGAAAGACAGATGAAATATTC.. 423
        |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
151  Leu11eProLysSerSerAsn11eGlySerSerLysTyrLysProVal1e 167
    424  ....TCACTTGACAGAGACAAAGCGAGAGACGATGTGTGATCAAAGA 469

```



```

66 AAAAAAAAAAAAAAAAAA.....AAACCGGGGGA 97
   |||
96 sthrvalthrproalalysalavalthrthrprogllyslsglyalatr 113
   |||
98 AACCGGGGAC...ATACGAGTCCCGGGCTGGAA...CTGGTAC 141
   |||
113 hrprogllyslaleuvalalatrhrprogllyslsglyalalatl 129
   |||
142 CCGGTACAAACCTCCAAAGAGAAACATTGTCAAGCGCTCCACTAC 191
   |||
130 proalalysgllyalalysasmelyllysasnalalysgluasps 146
   |||
192 GCGACACAGCTAGAGAAACACATCTACTAGAACCTCAACATATTAAC 241
   |||
146 poluglugluaspaspaspsercglugluaspclugluaspasp 163
   |||
242 ATAGTACAGAGAAACGCCACATAGACTACTGCATTAAGTCAACTGCG 291
   |||
163 luaspgluaspcgluaspcglutllecgluoproalaalmetlslala 179
   |||
292 CAAGACACGATTCAGACTGAGAGAGAAAGAACGACGACAGAGAGA 341
   |||
180 AlaAlaIaProAlaSerGlu...AspGluaspaspGluaspasp 195
   |||
342 CAACCAACGACGCGTACGCGGATGAGCAGACGCGCATACCCGA 391
   |||
195 poluaspsaspaspaspaspglugluaspaspsercglugluam 212
   |||
392 AAGAAACG.....GGGAAAGACAGATGAAGTAATCTCACTT... 429
   |||
212 luthrthrproalalysgllyslslalalalysvalalaproval 228
   |||
430 .....GACGAAGCAACCAAGCAGA 449
   |||
229 Alalysasnvalalagluaspclugluaspclugluaspcl 245
   |||
450 AGACGATGTTGATCAAGAAACACGCGGACGACAGAC 486
   |||
245 uaspaspaspaspaspaspgluaspcgluaspsasp 257

seq_name: p1r2:JE0378

seq_documentation_block:
DNA (cytosine-5)-methyltransferase (EC 2.1.1.37) - rat
C:Species: Rattus norvegicus (Norway rat)
C>Date: 23-Jul-1999 #sequence_revision 23-Jul-1999 #text_change 21-Jul-2000
C:Accession: JE0378
R:Kimura, H.; Takeda, T.; Tanaka, S.; Ogawa, T.; Shiota, K.
Biochem. Biophys. Res. Commun. 253, 495-501, 1998
A:Title: Expression of rat DNA (cytosine-5) methyltransferase (DNA MTase) in rodent trop
A:Reference number: JE0378; MUID:99097263
A:Accession: JE0378
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1622 <KIM>
A:Cross-references: DDBJ:AB012214; NID:q4160669; PIDN:BA37118.1; PID:q4160670
C:Keywords: methyltransferase; S-adenosylmethionine

```

```

alignment_scores:
  Quality: 96.50      Length: 149
  Ratio: 1.109      Gaps: 10
  Percent Similarity: 58.389      Percent Identity: 29.530

```

alignment_block:

US-09-823-101-4 x JE0378 ..

Align seg 1/1 to: JE0378 from: 1 to: 1622

```

57 GAGGAAAAAAAAAAAAAAAAAAAAAAAAAACCAGGGG...GAACACAG 103
   |||
191 Gtugluarqaspcluasplyslarsarqvalalaglythrcluserar 207

```

```

104 G.....GACATAGC...GAGTCCCGGGCTGGAACTGTGC 138
   |||
207 galaserarqalaglygluservalglulyspro.....gluarqvala 222
   |||
139 TACCCGCTACAAACCTCCAAAGAGAAACATTGTCAAGCGCTCCACT 188
   |||
222 rgprogllythrclneucysgclngclugluin...glyglugluasp 237
   |||
189 ACCGGCACACAGCTAGAGAAACACATCTACTAGAAACCTCAACATTA 238
   |||
238 asparqarqproarqarqglthrarqglueualaserarqarqlys.. 253
   |||
239 TACATAGTACAGAGAAACGCGACATAGAGCTACTGCATTAAGTCAACT 288
   |||
254 .....serarqgluasproasparqglualarqpro.glythr... 266
   |||
289 GCGCAAGACACGCTATCAGAGACTGACGAGAGAGAAACGACGACAGAG 338
   |||
267 .....HslleuAspvAlaspaspaspaspclutlslsAsplysArqS 280
   |||
339 AGACAACCAACGACGCGCGATGAGCGGATGAGCGACGACGACATAC 388
   |||
280 rserarqproarqserglu...ProarqaspLeuAlathrlslsArqArq 296
   |||
389 CGAAAGAAACGCGGAAAGACAGATGAAGTAATCTCACTTGCACGAC 438
   |||
296 rOlysglu.....gluvalglugluinlthrthrproglurpro 308
   |||
439 CAAGACGAGAACGACGATGATCAAGAAACACGCGGACAGA 483
   |||
309 gluglylslsAspGluaspgluarqglugluysarqarqlysthr 323

seq_name: p1r2:T19201

seq_documentation_block:
hypothetical protein C11G6.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C:Accession: T19201
R:Wilkinson, J.
Submitted to the EMBL Data Library, March 1996
A:Reference number: Z19089
A:Accession: T19201
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-385 <WIL>
A:Cross-references: EMBL:Z70204; PIDN:CAA94113.1; GSPR3:GN00028; CESP:C11G6.3
A:Experimental source: clone C11G6
C:Genetics:
A:Gene: CESP:C11G6.3
A:Map position: X
A:introns: 12/1; 106/3; 171/2; 335/2

```

```

alignment_scores:
  Quality: 96.00      Length: 137
  Ratio: 1.247      Gaps: 4
  Percent Similarity: 56.204      Percent Identity: 21.898

```

alignment_block:

US-09-823-101-4 x T19201 ..

Align seg 1/1 to: T19201 from: 1 to: 385

```

76 AAAAAAAAAAAACCGGGGGAACACGAGGACATAGCGAGTGTCCGGG 125
   |||
87 lylslsglugluuproserSerSerSerSerSerSerSerSerSer 102
   |||
126 CTGGAACCTGTGTACCGCGGTACAAACCTCCAAAGAGGAACATTTGC 175
   |||
102 ySerSerleugluthrthrproSerSerSerSerSerSerSerSer 119

```

```

176 AACGGCTCCACTACCGCGACACGCTAGAGAAACACATCTACTAGAAA 225
    ::::  |||  ::  |||
119 ISHISHSLysLysGluArgLysAspLysGlu..... 129
226 CTCACACATATATACATAGTACAGAGAAAGCCACATAGACTCTCGC 275
    ||:::  :::::  |||
130 .....HisLysLysHisLysLysAspArgGlu..... 138
276 ATAAAGTAACTGCGCAAGACGATATCAGAGACTGACGAGAGAA 325
    |||  :::::  |||||
139 .....HisArgAspArgGluLysGluArgGlu 148
326 ACGACGACAGAGAGACAAACCAAGCCAG.....CGTAGCCGCAT 369
    ||:::  :::::  |||||
148 rgsaprgluArgLysgluArggluArggluLysgluLysgluArgglu 164
370 GAGCGACAGCGCGAGATACCGCAAGAAACGGGCAAAAGACAGATGAAGT 419
    :::::  |||||  :::::  |||
165 AspAlaAlaArgArgGluIleGluGluLysAlaGluMetAspAlaLysAr 181
420 AATCTCAGTTGACGAGAACCAAGACGAGAGACGATGTCATCAAGAA 469
    :::::  :::::  |||||
181 gValAlaGluGluGluGluArgArgLysGluLysGluLysArgArg 198
470 ACCAGCCGAG 480
    :::::
198 IuGluLysLys 201

```

seq_name: p1r2:T21435

```

seq_documentation_block:
hypothetical protein F26h11.3c - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 02-Sep-2000
C:Accession: T21435
R:Barlow, K.
submitted to the EMBL Data Library, November 1996
A:Reference number: Z19421
A:Accession: T21435
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-452 <Wt>
A:Cross-References: EMBL:Z81515; PIDN:CAB54234.1; GSPDB:GN00020; CESP:F26H11.3c
A:Experimental source: clone F26h11
C:Genetics:
A:Gene: CESP:F26H11.3c
A:Map position: 2
A:Introns: 58/1; 89/3; 230/2; 290/2; 334/1
C:Superfamily: bromodomain homology
F:313-368/Domain: bromodomain homology <BRO>

```

```

alignment_scores:
  Quality: 96.00      Length: 118
  Ratio: 1.371      Gaps: 5
Percent Similarity: 59.322      Percent Identity: 27.966

```

alignment_block:
us-09-823-101-4 x T21435 ..

```

Align seg 1/1 to: T21435 from: 1 to: 452
63 AAAAAAAAAAAAAAAAAAAACCCGGGAGAAACGAGGACATAG 112
    |||||:  :::::  |||||  |||||
24 LysLysGluLysLysArgArgThrArgMetLysngluAlaGlu 40
113 CGAGTCCCGCGGCTGGAACCTGTGTACCCGCTACAAACCTCAAAAG 162
    :::::  |||  |||  |||
40 ngLysGluLysLeuLeuThrPro.Trp.....ArgLysAlaArgAla 53
163 AGGAAATATGTCAACGCGCTCCACTACCGCGCACACAGCTAGAGAAACA 212
    |||  |||  |||||  |||
54 ArgProHis.....ArgAlaAlaLysProLysAlaGluValLysLysGlu 68

```

```

213 CATCTACTAGAAACTC.....ACAACATATATACATA 244
    :::  |||||  :::
68 uValIleAsnProLysAspIleThrLeuGlyGlyAspThrTyTrAspTyr 85
245 GTACAGAGAAACGCCACATAGAGCTACTCGCATTAAGTACAACTGCGCAA 294
    ||:::  |||  :::::  |||
85 allYsgluGluLysProThrGluSerIleAlaThrAsnValSerArg 101
295 AGACACGTATCAGAGACTGACGAGAGAGAAACGACGACGAGAGACAA 344
    ||:::  |||  :::::  |||||
102 ArgArgThrSerIleAsnLeuSerLysSerGluAspArg...AspLys 117
345 ACCAAGCAGCCAGCGCTAGCCGCGATGAGCGACGACGCGATACCGCAAG 394
    ||||  |||||
117 sProGluSerGluSerThrAlaProLysSerLysGluArgArgThrSer 134
395 AA 396
    ||
134 Iu 134

```

seq_name: p1r2:T46481

```

seq_documentation_block:
hypothetical protein DKFZp434A025.1 - human (fragment)
C:Species: Homo sapiens (man)
C>Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 04-Feb-2000
C:Accession: T46481
R:Ducresthoelt, A.; Lauber, J.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, January 2000
A:Reference number: Z23035
A:Accession: T46481
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1027 <AAA>
A:Cross-References: EMBL:AL137755
A:Experimental source: adult testis; clone DKFZp434A025
C:Genetics:
A:Note: DKFZp434A025.1

```

```

alignment_scores:
  Quality: 96.00      Length: 174
  Ratio: 1.043      Gaps: 5
Percent Similarity: 52.874      Percent Identity: 21.839

```

alignment_block:
us-09-823-101-4 x T46481 ..

Align seg 1/1 to: T46481 from: 1 to: 1027

```

43 CGTAAATCTATCAAGAGAAAAA..... 84
    :::::  |||  |||  :::::  |||||
42 LysAspHisIleAspArgThrArgLysLysArgGlyGluLysAspGlu 58
85 .....AAAC 88
58 rGluTyrGluTyrSerGlySerGluGluGluGluGluValProGlu 75
89 CCGGGGGGAAACGAGGACATAGCGAGTGTCCGGGCTGGAAA...CTG 135
    ||:::  ||:::  ||:::  |||||
75 IuGluGluGluProSerSerIleValAsnValProGluGluSerThrLeu 91
136 TGGTACCCGGCTACAAACCTCCAAAGAGGAAACATTGTCAAGCGCGCTCC 185
    :::::  |||||
92 ArgArgAspPheLeuArgLeuGluGluAsnLysLysLysArgSerGlu 108
186 ACTACCGCGCACACGCTAGAGAAACATCTACTAGAAACATCAACAAT 235
    |||  |||  |||||  :::::
108 aleuArgArgGluGluLeuGluGlu.....GluG 119
236 ATATACATAGTACAGAGAAACGCCACATAGAGCTACTCGCA..... 276
    :::::  |||||

```



```

119  lnteuatrgtuglntuglntuyrlytsatrglntleuenaiaagluatrgln 135
277  ..... TAAAGTRCAACTGCCAAGACACATGACAGACTCA 314
136  lvsatrglleglntuglntlntylsuglntnatrgatrgatrgleuuglntugln 152
315  CGAGACACAAACAGCAGCAGAGAGAGAACAAACCAACAGCAGCGTAGC 364
152  natrgatrgluatrgluatlaatrgatrglntuglnlntuarglntnatrgatrga 169
365  GCATGATGACACAGCGCGACATCCGAAAGAAACGGGGAAAGACAGATG 414
169  rgtuglntuglntuglntuytsatrgatrgleuglntuglntleuuglntatrgatrgatrglys 185
415  AAGGTAATCTCAAGTTTCACGACCAACCAAGACAGACAGATGTTGTATCA 464
186  glntu..... glntuglntuargatrgatrglaaglntuglntuytsatrg 198
465  AAGCAACCAAGCGCGACAGACAG 486
198  gatrgvalatrgatrglntuglntu 205

```

THIS PAGE BLANK (USPTO)

OM of: US-09-823-101-4 to: SwissProt_39.* out-format : pfs

Date: Jan 17, 2002 4:08 PM

About: Results were produced by the Gencore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL-frame+ n2p.model -DEV-xlp
-O-qgn2.1/USPTO.spool/US09823101/runat_17012002.155138.22061/app_query.fasta_1.2239
-DB-SwissProt_39 -OPMT-fastan -SUFFIX-n2p.rsp -GAPOP-12.000
-GAPEXT-4.000 -MINMATCH-0.100 -LOOPEL-0.000 -LOOPEXT-0.000
-QGAPOP-4.500 -QGAPEXT-0.050 -XGAPOP-10.000 -XGAPEXT-0.500
-FGAPOP-6.000 -FGAPEXT-7.000 -YGAPOP-10.000 -YGAPEXT-0.500
-DELOP-6.000 -DELEXT-7.000 -START-1 -MATRIX-biosum62
-TRANS-human40.cdi -LIST-45 -DOCALIGN-200 -THR SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN-15 -MODE-LOCAL -OUTFMT-pfs
-NORM-ext -MINLEN=0 -MAXLEN=200000000
-USPT-US09823101.qcgn1_1_0 -NCPU-6 -ICPU-3 -LONGLOG -NO_XLPXY
-MAIT -THREADS=1

Search information block:

Query: US-09-823-101-4
Query length: 509
Database: SwissProt_39.*
Database sequences: 100059
Database length: 36664827
Search time (sec): 66.470000

score_list:

Sequence	Stid	Orig	Zscore	EScore	Len	Documentation
SwissProt_39:MTDM_MOUSE	110	00	166.14	0.0086	1620	P13864 mus musculus (mouse)
SwissProt_39:YKR2_YEAST	97	00	156.86	0.1380	427	P36049 saccharomyces cerevisiae
SwissProt_39:NUCL_HUMAN	96	50	150.03	0.1562	706	P19338 homo sapiens (human)
SwissProt_39:XPR6_YARLI	95	00	144.94	0.2170	976	P42781 yarrowia lipolytica
SwissProt_39:TFHY_SHEEP	94	50	140.44	0.2437	1549	P22793 ovis aries (sheep)
SwissProt_39:TD3_HUMAN	94	00	142.43	0.2697	1083	O00068 homo sapiens (human)
SwissProt_39:SNL1_YEAST	94	00	136.70	0.2731	2231	O00416 saccharomyces cerevisiae
SwissProt_39:SPR8_HUMAN	92	00	145.46	0.4094	484	O05519 homo sapiens (human)
SwissProt_39:NUCL_MESAU	91	50	141.55	0.4590	713	P08199 mesocricetus auratus
SwissProt_39:GAG_MSVMO	91	00	142.94	0.5088	538	P03334 moloney murine sarcoma
SwissProt_39:INCE_CHICK	91	00	139.07	0.5131	877	P53352 gallus gallus (chicken)
SwissProt_39:GAG_MLVMO	90	50	142.07	0.5668	540	P21435 homi murine leukemia
SwissProt_39:YUJ2_CAEEL	90	00	145.79	0.6250	304	P33864 caenorhabditis elegans
SwissProt_39:ATRX_HUMAN	89	00	127.80	0.8036	2375	P46100 homo sapiens (human)
SwissProt_39:NM52_DRBY	88	50	131.20	0.8868	1391	O08096 drosophila hydei (fruit fly)
SwissProt_39:NM54_TRYBB	88	00	137.17	0.9731	590	P04540 trypanosoma brucei (tsetse fly)
SwissProt_39:IF3A_MALIZ	88	00	133.29	0.9814	962	O9xh2 zea mays (maize) . euk
SwissProt_39:SPR8_HUMAN	87	50	133.54	1.09	951	O12872 homo sapiens (human)
SwissProt_39:CCAB_DDISM	87	50	125.44	1.11	2326	P56698 discipye ornata (cel)
SwissProt_39:NMN4_YEAST	87	00	130.00	1.22	1178	P36044 saccharomyces cerevisiae
SwissProt_39:IF2P_HUMAN	87	00	129.72	1.22	1220	O06041 homo sapiens (human)
SwissProt_39:ASA_EHENT	87	00	127.20	1.23	1676	P35269 emericella nidulans
SwissProt_39:TF2A_HUMAN	86	50	135.69	1.34	517	P35269 homo sapiens (human)
SwissProt_39:GAG_MLVMO	86	50	135.38	1.34	538	P03332 moloney murine leukemia
SwissProt_39:GAGA_SOYBN	85	50	133.76	1.67	605	P13916 glycine max (soybean)
SwissProt_39:TFHY_RAITT	85	50	126.07	1.69	1407	P37709 cryctolagus cuniculi
SwissProt_39:HRX_MOUSE	85	50	118.05	1.72	3866	P55200 mus musculus (mouse)
SwissProt_39:K2_DROME	85	00	126.54	1.88	1192	O46072 drosophila melanogaster
SwissProt_39:MYO6_MOUSE	85	00	126.07	1.89	1265	O64331 mus musculus (mouse)
SwissProt_39:SON_HUMAN	85	00	124.60	1.89	1523	P18583 homo sapiens (human)
SwissProt_39:GAG_MLVMO	84	50	132.04	2.07	536	P29168 murine leukemia virus
SwissProt_39:GAG_MLVMO	84	50	132.04	2.07	539	P26807 friend murine leukemia
SwissProt_39:RDI_PIG	84	50	131.38	2.07	583	P26044 sus scrofa (pig) . radi
SwissProt_39:SC1_RAT	84	50	130.21	2.07	675	O35314 ratus norvegicus (rat)
SwissProt_39:GLYB_PLAFG	84	50	129.13	2.08	774	P02895 plasmodium falciparum
SwissProt_39:TOP1_XENLA	84	50	128.55	2.08	829	P41512 xenopus laevis (afric)
SwissProt_39:YUJ1_YEAST	84	00	124.55	2.34	1240	P53335 saccharomyces cerevisiae
SwissProt_39:MYO6_HUMAN	84	00	124.41	2.34	1262	O9um54 homo sapiens (human)
SwissProt_39:SEC7_YEAST	83	50	119.88	2.62	2009	P11075 saccharomyces cerevisiae
SwissProt_39:TOD2_CAEEL	83	00	136.13	2.81	233	P34594 caenorhabditis elegans

SwissProt_39:IF3A_TODAC + 83.00 124.92 2.88 958 O40554 nicotiana tabacum
SwissProt_39:GLT_DROME + 83.00 124.39 2.89 1023 P3438 drosophila melanog
SwissProt_39:HRP1_SCHPO + 83.00 122.06 2.90 1373 O9us25 schizosaccharomyc
SwissProt_39:CG7_HUMAN + 82.50 132.04 3.16 351 O9y388 homo sapiens (huma
SwissProt_39:CBF5_KULIA + 82.50 129.66 3.17 474 O13473 kluyveromyces lact

seq_name: SwissProt_39:MTDM_MOUSE

seq_documentation_block:
ID MTDM_MOUSE STANDARD; PRT; 1620 AA.

AC P13864:
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE DNA (CYTOSINE-5)-METHYLTRANSFERASE MMU1 (EC 2.1.1.37) (DNA
DE METHYLTRANSFERASE MMU1) (DNA MTASE MMU1) (MCMU) (M.MMU1).
GN DNMT1 OR DNMT OR UDM.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
[1]
RX SEQUENCE FROM N.A.
RX MEDLINE=69094873; Pubmed=3210246;
RA Bestor T.H., Laudano A., Mattaliano R., Ingram V.;
RT "Cloning and sequencing of a cDNA encoding DNA methyltransferase of
RT mouse cells. The carboxyl-terminal domain of the mammalian enzymes is
RT related to bacterial restriction methyltransferases.";
RL J. Mol. Biol. 203:971-983(1988).
[2]
RP REVISIONS TO N-TERMINUS.
RC TISSUE=Embryo;
RX MEDLINE=97094871; Pubmed=8940105;
RA Yoder J.A., Yen R.-W.C., Vertino P.M., Bestor T.H., Bayliss S.B.;
RT "New 5' regions of the murine and human genes for DNA (cytosine-5)-
RT methyltransferase.";
RL J. Biol. Chem. 271:31092-31097(1996).
CC -1- FUNCTION: METHYLATES CG RESIDUES.
CC -1- CATALYTIC ACTIVITY: S-ADENOSYL-L-METHIONINE + DNA -> S-ADENOSYL-L-
CC HOMOCYSTEINE + DNA CONTAINING 5-METHYLCYTOSINE.
CC -1- SIMILARITY: HIGH, TO OTHER EUKARYOTIC DNA METASE.
CC -1- SIMILARITY: SOME TO BACTERIAL RESTRICTION SYSTEMS.
CC METHYLTRANSFERASES.
CC -1- SIMILARITY: CONTAINS 2 BAH DOMAINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/annouce/>
CC or send an email to license@isb-sib.ch).

CC EMBL: X14805; CAJ32910.1; .
CC PIR: S01845; S01845.
CC REBASE: 2844; M.Mmu1.
CC MGD: MGI:94912; Dnmt1.
CC InterPro: IPR001025; BAH.
CC InterPro: IPR001525; C5 DNA_meth.
CC InterPro: IPR002857; ZNF-CXXC.
CC Pfam: PF01426; BAH; 2.
CC Pfam: PF00145; DNA_methylase; 3.
CC Pfam: PF02008; ZF-CXXC; 1.
CC PRINTS: PR00105; C5METRFRASE.
CC SMART: SM00439; BAH; 2.
CC PROSITE: PS00094; C5_MTASE_1; 1.
CC PROSITE: PS00095; C5_MTASE_2; 1.
CC TRANSFERASE: Methyltransferase; DNA-binding.
CC DOMAIN 651 693 CYS/ARG/LYS-RICH.
CC BY SIMILARITY.
CC AC SITE 1229 1229
CC SEQUENCE 1620 AA; 183286 MW; F73710AD043E709 CRC64;

alignment_scores: Quality: 110.00 Length: 148
 Ratio: 1.325 Gaps: 7
 Percent Similarity: 56.081 Percent Identity: 28.378

alignment_block:

US-09-823-101-4 x MTDN_MOUSE ..

Align seg 1/1 to: MTDN_MOUSE from: 1 to: 1620

```

54 CAAGAGGAGAAAAA.....AAAAAAAAA.....CCGGGGGGAACAG 103
   :::::..... :::::..... :::::..... :::::.....
192 GluGluThrGlnAspGlnAspLysArgValAlaValAspThrGlnSerG1 208
104 GGGACATAGCGAGTGTCCCGGCTGGAACCTGTGTACCGCGCTAC.... 149
   :::::..... :::::..... :::::..... :::::.....
208 YAlaAlaAlaValGluLysLeuGluGluValAlaThrAlaGlyThrGln 225
150 .....AAACTCCAAAGAGGAACATTGTCAACGCGCTCCACTAC 191
   :::::..... :::::..... :::::..... :::::.....
225 euGlyProGluGluProCysGluGluGlnLysAspAsnArgSerLeuArg 241
192 GCGGACACACTAGAGAAACACATCTACTAGAAACTCACAACATATATAC 241
   :::::..... :::::..... :::::..... :::::.....
242 ArgHisThr...ArgGluLeuSerLeuArgArgLysSer..... 253
242 ATAGTACAGAGAAACGCCACATAGACTACTCGCATAAAGTACAACTGCG 291
   :::::..... :::::..... :::::..... :::::.....
254 .....LysGluAspProAspArgGluAlaArgProGluThr..... 265
292 CAAGACACACTATCAGAGACTGACGAGAGAAGAAGCAGACAGAGAGA 341
   :::::..... :::::..... :::::..... :::::.....
266 .....HisLeuAspGlnAspGlnAspGlyLysLysAspLysArgSer 280
342 CAAGACACAGCGCAGCGCTGAGCGCATGAGCGAGCGCAGATACCCGA 391
   :::::..... :::::..... :::::..... :::::.....
280 ArgProArgSerGln...ProArgAspProAlaAlaLysArgArgPro 296
392 AAGAAGCGGGGAAAAAGACAGATGAAGTAATCTCAGTTGACGAAGACAA 441
   :::::..... :::::..... :::::..... :::::.....
296 YsgLysValArgAlaArgAlaGlySer.....SerArgAspSerGluAsp 310
442 AAGCGAGAGACGATGTTGATCAAGAAACGAGCGGAGACA 483
   :::::..... :::::..... :::::..... :::::.....
311 ArgAspGlnAspGlnArgGluGluLysArgArgLysThrThr 324

```

seq_name: SwissProt_39:YKR2_YEAST

seq_documentation_block:

```

ID YKR2_YEAST       STANDARD:       PRT:   427 AA.
AC P36049;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE HYPOTHELTICAL 49.7 KDA PROTEIN IN GIN2-STE3 INTERGENIC REGION.
GN YKL172W OR YKL636.
OS Saccharomyces cerevisiae (Baker's Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RX MEDLINE=94378719; PubMed=8091858;
RA Vandenbol M., Bolle P.-A., Dion C., Portetelle D., Hilger F.;
RT "Sequencing and analysis of a 20.5 kb DNA segment located on the left
   arm of yeast chromosome XI.";
RL Yeast 10:525-533(1994).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```

CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

```

-----
DR EMBL: Z26878; CA81515.1; -.
DR EMBL: Z28172; CA82014.1; -.
DR PIR: S38002; S38002.
DR PIR: S38409; S38409.
DR PIR: S44589; S44589.
DR SGD: S0001655; YKL172W.
KW Hypothetical protein.
SQ
SEQUENCE 427 AA; 49734 MW; 4A11F6CDE779DB5A CRC64;

```

alignment_scores: Quality: 97.00 Length: 161
 Ratio: 1.102 Gaps: 9
 Percent Similarity: 54.658 Percent Identity: 24.845

alignment_block:

US-09-823-101-4 x YKR2_YEAST ..

Align seg 1/1 to: YKR2_YEAST from: 1 to: 427

```

37 AAAAAAGTAAATCTATCAAGAGAAAAA.....AAAAAAA 83
   :::::..... :::::..... :::::..... :::::.....
22 LysLeuGlnLysAsnLysLeuLysLysSerGlnGluLeuLysLysG1 38
84 AAAACCCGGGGGAAACCAAGGACATAGCGACTGTCCCGGCTGGAAC 133
   :::::..... :::::..... :::::..... :::::.....
38 uGluPro.....ThrIleValThrAlaSerAsnLeuLys. 49
134 TGGTGTACCCGGCTACAAACTCCAAAGAGGAACATTGTCAACGGCT 183
   :::::..... :::::..... :::::..... :::::.....
50 .....LysLeuGluLysGlu..... 55
184 CCACTACCCGGCACACAGCTAGAGAAACACATCTACTAGAACTACAAC 233
   :::::..... :::::..... :::::..... :::::.....
56 .....LysLysAlaAspValLysLysGluValAlaAlaAspThrGlu 70
234 ATAT.....ATACATAGTACAGAGAAACCCACATA.....G 265
   :::::..... :::::..... :::::..... :::::.....
70 uTyrGlnSerGlnAlaLeuSerLysLysGluLysArgLysLeuLys 87
266 AGCTATCTC.....GCATTAAGTACACTGCGCAAGACACGTA 303
   :::::..... :::::..... :::::..... :::::.....
87 LysLeuLysLysMetGlnGluGlnAspAlaThrGluAlaGlnLysHisMet 103
110 rGlyAspAspArgGluGluGlnGluGluGluGluGluGluGly 127
404 AAAGACAGTGAAGTAATCTCAGTTGACGAGAACCAAGCAAGAAC 453
   :::::..... :::::..... :::::..... :::::.....
127 rGluAspLysLeuGluLysLeuAla...LysSerSerGlnSerGluAsp 142
454 GATGTTGATCAAGAAACGAGCGAGACGAGAC 486
   :::::..... :::::..... :::::..... :::::.....
143 AspSerGlnSerLysAsnAspSerGlnGluAsp 153

```

seq_name: SwissProt_39:NUCL_HUMAN

seq_documentation_block:

```

ID NUCL_HUMAN       STANDARD:       PRT:   706 AA.
AC P19338;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE NUCLEOLIN (PROTEIN C23).

```

NCL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-89290043; PubMed-2737305;
RA Strivastava M., Fleming P.J., Pollard H.B., Burns A.L.;
RT "Cloning and sequencing of the human nucleolin cDNA";
RL FEBS Lett. 250:99-105(1989).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE-90368666; PubMed-2394707;
RA Srivastava M., McBride O.W., Fleming P.J., Pollard H.B., Burns A.L.;
RT "Genomic organization and chromosomal localization of the human
nucleolin gene";
RL J. Biol. Chem. 265:14922-14931(1990).
CC -! FUNCTION: NUCLEOLIN IS THE MAJOR NUCLEAR PROTEIN OF GROWING
EUKARYOTIC CELLS. IT IS FOUND ASSOCIATED WITH INTRANUCLEAR
CHROMATIN AND PRERIBOSOMAL PARTICLES. IT INDUCES CHROMATIN
DECONDENSATION BY BINDING TO HISTONE H1. IT IS THOUGHT TO PLAY A
ROLE IN PRE-RRNA TRANSCRIPTION AND RIBOSOME ASSEMBLY.
CC -! SUBCELLULAR LOCATION: NUCLEAR; NUCLEOLS.
CC -! SIMILARITY: CONTAINS 4 RNA RECOGNITION MOTIFS (RRM).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed, usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

DR EMBL: M60858; AAA59954.1; .
DR PIR: A35804; A35804.
DR PIR: S04631; S04631.
DR HSSP: P09651; IUP1.
DR Aarhus/Ghent-2DPAGE; 1210; NEPHGE.
DR MIW: 164035; .
DR InterPro; IPR000504; RRM.
DR Pfam; PF00076; rrm; 4.
DR SMART: SM00360; RRM; 4.
DR PROSITE; PSS0102; RRM; 4.
DR PROSITE; PS00030; RRM_RNP_1; 3.
KW Nuclear protein; phosphorylation; Methylation; DNA-binding; Repeat;
KW RNA-binding.
FT INT MET 0
FT DOMAIN 142 170 ASP/GLU-RICH (ACIDIC).
FT DOMAIN 184 208 ASP/GLU-RICH (ACIDIC).
FT DOMAIN 233 270 ASP/GLU-RICH (ACIDIC).
FT DOMAIN 306 382 RNA-BINDING (RRM) 1.
FT DOMAIN 392 465 RNA-BINDING (RRM) 2.
FT DOMAIN 485 559 RNA-BINDING (RRM) 3.
FT DOMAIN 571 643 RNA-BINDING (RRM) 4.
FT DOMAIN 645 694 ARG/GLU/PHE-RICH.
FT DOMAIN 57 134 8 X 8 AA TANDEM REPEATS OF X-T-P-X-K-K-
X-X.
FT REPEAT 57 64 1.
FT REPEAT 74 81 2.
FT REPEAT 82 89 3.
FT REPEAT 90 97 4.
FT REPEAT 97 103 5.
FT REPEAT 104 111 6.
FT REPEAT 119 126 7.
FT REPEAT 127 134 8.
FT MOD_RES 144 144 PHOSPHORYLATION (BY SIMILARITY).
FT MOD_RES 152 152 PHOSPHORYLATION (BY SIMILARITY).
FT MOD_RES 183 183 PHOSPHORYLATION (BY SIMILARITY).
MOD_RES 706 AA; 76213 MW; 85A2FCA22EA03DB CQC64;
SEQUENCE

Quality:	96.50	Length:	179
Ratio:	1.027	Gaps:	6
Percent Similarity:	52.514	Percent Identity:	24.022
alignment_block:			
US-09-823-101-4	x	NUCL_HUMAN	..
Align seg 1/1	to:	NUCL_HUMAN	from: 1. to: 706

Align seq 1/1 to: NUCL_HUMAN from: 1. to: 706

```

16 AGCATGCTGATATCATTCAGGAGAAAAACGTAATTCATCAAGAGGAAAA 65
   ::::::::::::::| | | | | | | | | | | | | | | | | | | | | |
79 LysLaLaLaValThrProGlyLysLaLaLaLaLaLaThrProAlaLys 95
   AAAAAAAAAAAAAAAAAA.....AAACCCGGGGGA 97
   | | | | | | | | | | | | | | | | | | | | | |
95 sThnValThrProAlaLysAlaValThrThrProGlyLysLysGlyAlaT 112
   AACCAAGGGGAC...ATACGAGTGTCCGGGCTGAAA...CTGTGTAC 141
   | | | | | | | | | | | | | | | | | | | | | |
112 hrProGlyLysAlaLeuValAlaThrProGlyLysGlyAlaAlaLe 128
   CCGGCTCAAAACTCCAAAAGAGAAACCTGTCACAGCGGCCACATCC 191
   | | | | | | | | | | | | | | | | | | | | | |
129 ProAlaLysGlyAlaLysasnGlyLysAsnAlaLysLysGlyAspSer 145
   GCGCACACACCTAGAGAAACACATCTACTACAGAACTCAACATATAT 192
   ::::: | | | | | | | | | | | | | | | | | | | | | |
145 pGluGluGluAspAspSerSerGluGluAspGluAspGluAspGlu 162
   ATATGACAGAGAAAGCCGACATAGACTACTGCGCTAAAGTACAACTGC 291
   ::::: | | | | | | | | | | | | | | | | | | | | | |
162 LuAspGluAspGluAspGluLeuGluProAlaAlaMetLysLaLaLa 178
   CAAAGACACCTATCAGAGACTGACGAGAGAAAGAACGACAGAGAGA 341
   ::::: | | | | | | | | | | | | | | | | | | | | | |
179 AlaAlaThrProAlaSerGlu...AspGluAspAspGluAspAspGlu 194
   CAAACCAAGCAGCCAGCGTGAAGCGCGATGAGCGACGACGATCCCGA 391
   ::::: | | | | | | | | | | | | | | | | | | | | | |
194 pGluAspAspAspAspAspGluGluAspSerSerGluGluAlaMet 211
   AAGAAAGC.....GGAAAGACACATGAAGTATCTCAGTT... 429
   :: | | | | | | | | | | | | | | | | | | | | | |
211 LuThrThrProAlaLysGlyLysLysAlaAlaLysAlaValProVal 227
   430 .....GACGAAGACCAAAAGCGAGA 449
   ::::: | | | | | | | | | | | | | | | | | | | | | |
228 AlaLysAsnValAlaGluAspGluAspGluGluGluAspAspGluAsp 244
   450 AGACGATGTCATCAAAAGAAACACGCGCCAGACACAC 486
   | | | | | | | | | | | | | | | | | | | | | |
244 uAspAspAspAspAspGluAspGluAspGluAspAsp 256

seq_name: SwissProt_39:XPR6_YARLI

seq_documentation_block:
ID XPR6_YARLI STANDARD: PRT; 976 AA.
AC P42781;
DC 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE DIBASIC PROCESSING ENDOPROTEASE PRECURSOR (EC 3.4.21.-).
GN XPR6.
OS Yarrowia lipolytica (Candida lipolytica).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Dipodascaceae; Yarrowia.
OX NCBITaxID=4952;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CX161-1B;
RX MEDLINE=94262316; PubMed=8203153;
RA 'Endlin C.S., Ogrzyzlak D.M.'
"Cloning, nucleotide sequence and functions of XPR6, which codes for

```

```

Align seg 1/1 to: XPr6_YARLti from: 1 to: 976

61 AAAAAAAAAAAAAAAAAAAAAACCCGGGGGCAACCCAGGGGACAT 110
   |||:::|||||:::|||||:::|||||:::|||||:::
103 ArgLysArgGlnLysLeuGlnLysTyrGlyMetSerGlnLysAsp 119
   |||:::|||||:::|||||:::|||||:::|||||:::
111 AGCGAGT.....CTCCCGGGCTGGAAACGTGGTCCCGCGCATCAAAA 153
   :::::|||||:::|||||:::|||||:::|||||:::
119 GlnLysArgLeuValAlaLeuGlnLysAspTyrAspTyrSerGln 136
   :::::|||||:::|||||:::|||||:::|||||:::
154 .....CTCCAAAGAGGAAACATWTGTCAACGC 180
   :::::|||||:::|||||:::|||||:::|||||:::
136 rgGlyLeuGlySerLeuGlnValLeuSerGlnLysArgGlnIleHisLysArg 152
   :::::|||||:::|||||:::|||||:::|||||:::
181 GCTCCACTACCGCGCACACAGCTAGAGAAACACATCTACTAGAAACTCAC 230
   |||:::|||||:::|||||:::|||||:::|||||:::
153 AlaProValAsnTyrPheGlnGluGluMetGlu..... 163
   |||:::|||||:::|||||:::|||||:::|||||:::
231 AACATATATACATAGTGTACAGGAAGCCCATATAGAGCTCTCGCATAAA 280
   |||:::|||||:::|||||:::|||||:::|||||:::

```

```

164 ...Tyrleuylsgulileylsatgaty..... 171
281 GTACAACTGCCGAAAGACACGTATCAGAGCTACGACAGAGAAGACGAC 330
172 .....AlaIugluAlaIugIntlysAlaIugInsp 180
331 GACAGAGGAGCAAAACCAACGAGCCAG.....CGTAGCGCCGA 368
181 AsplyrsglyaspIylsYsgIuaspIuInlyaspAsplyrsglysgIuglyl 197
369 TGGAGCAGACGAGCATACCCGAAAGAAACGGGAAAGACAGATGAAAG 418
197 ngluAlaIugIntlysgIugIyaspIylsAsplyrsglysgIuglyasp..... 211
419 TAATCTCAGTTGACGAGAACCAAAAGCGAGAGAGCATGTTGATCAAGA 468
212 .....AsplyrsgIuaspIyIugIugIuaspAspAspAspIu 224
469 AACGACGCCGACAGACTGTGCA 492
225 AspGIuaspAspAspAlaSer 232

seq_name: SwissProt_39:TRHY_SHEEP

seq_documentation_block:
ID TRHY_SHEEP STANDARD; PRT; 1549 AA.
AC P22793;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE TRICHOHYALIN.
GN THH.
OS Ovis aries (Sheep).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Ovis.
OX NCBI_TaxID=9940;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93260018; PubMed=7684041;
RA Fietz M.J., McLaughlan C.J., Campbell M.T., Rogers G.E.;
RT "Analysis of the sheep trichohyalin gene: potential structural and
RT calcium-binding roles of trichohyalin in the hair follicle.";
RL J. Cell Biol. 121:855-865(1993).
RN [2]
RP SEQUENCE OF 1016-1549 FROM N.A.
RX SRINAI-MERINO-DORSET HORN X BORDER LEICESTER; TISSUE=WOOL follicles;
RX MEDLINE=90130632; PubMed=2298812;
RA Fietz M.J., Presland R.B., Rogers G.E.;
RT "The cDNA-deduced amino acid sequence for trichohyalin, a
RT differentiation marker in the hair follicle, contains a 23 amino acid
RT repeat.";
RL J. Cell Biol. 110:427-436(1990).
RN [3]
RP FUNCTION: INTERMEDIATE FILAMENT-ASSOCIATED PROTEIN THAT ASSOCIATES
CC IN REGULAR ARRAYS WITH KERATIN INTERMEDIATE FILAMENTS (KIF) OF THE
CC INNER ROOT SHEATH CELLS OF THE HAIR FOLLICLE AND THE GRANULAR
CC LAYER OF THE EPIDERMIS. IT LATER BECOMES CROSS-LINKED TO KIF BY
CC ISOPEPTIDE BONDS. IT MAY SERVE AS SCAFFOLD PROTEIN, TOGETHER
CC WITH INVOLICRIN, IN THE ORGANIZATION OF THE CELL ENVELOPE OR EVEN
CC ANCHOR THE CELL ENVELOPE TO THE KIF NETWORK. IT MAY BE INVOLVED IN
CC ITS OWN CALCIUM-DEPENDENT POSTSYNTHETIC PROCESSING DURING TERMINAL
CC DIFFERENTIATION.
CC -1 SUBUNIT: HOMODIMER (PROBABLE).
CC -1 ALTERNATIVE PRODUCTS: AT LEAST TWO FORMS MAY BE PRODUCED BY
CC ALTERNATIVE SPLICING OF THE SAME GENE.
CC -1 TISSUE SPECIFICITY: FOUND IN THE HARD KERATINIZING TISSUES SUCH AS
CC THE INNER ROOT SHEATH (IRS) OF HAIR FOLLICLES AND MEDULLA, AND IN
CC THE EPITHELIA OF THE TONGUE, HOOF AND RIMEN.
CC -1 DOMAIN: CONSISTS OF NINE DOMAINS. DOMAIN 1 CONTAINS TWO EF-HAND
CC CALCIUM-BINDING DOMAINS. DOMAINS 2-4, 6, AND 8 ARE ALMOST
CC ENTIRELY ALPHA-HELICAL, CONFIGURED AS A SERIES OF PEPTIDE REPEATS
CC OF VARYING REGULARITY, AND ARE THOUGHT TO FORM A SINGLE-STRANDED
CC ALPHA-HELICAL ROD STABILIZED BY IONIC INTERACTIONS. DOMAIN 6 IS

```


SQ SEQUENCE 713 AA: 76997 MW: 79DDCF724CED7B4 CRC64;

alignment_scores:

Quality: 91.50 Length: 170
Ratio: 0.943 Gaps: 7
Percent Similarity: 57.059 Percent Identity: 24.118

alignment_block:
US-09-823-101-4 x NUCL_MESAU ..

Align seg 1/1 to: NUCL_MESAU from: 1 to: 713

```

38 AAAAAAGTAAATCTATCAAGAGAAAAA.....AA 69
|||||:|||||:|||||:|||||:
60 LysLysValValValSerGlnThrLysLysValAlaValProThrProAl 76
70 AAAAAAAAAAAAAACCGGGGGAACCGGAGCATGAGCATG 119
|||||:|||||:|||||:|||||:
76 aLysLysAlaAlaValThrProGlyLysLysAlaAlaAlaThrProAl 93
120 CCGGGGCGGAACCTGTGTACCGGGCTACAAACTCCA...AAAGAGA 166
|||||:|||||:|||||:|||||:
93 yslYsAla.ValThrProAlaLysAlaValAlaThrProGlyLysLysG 109
167 AACATTGTCACAGCGCTACCGGCGACAGCTAGAGAAACACATC 216
|||||:|||||:|||||:|||||:
109 yAlaThrGlnAlaLysAlaLeuValAlaThrProGlyLysLysGlyAla 126
217 TACTAGAAACTCACACATATATATAGTACAGAGAAACCGCACATAGA 266
|||:|||||:|||||:|||||:
126 A.....ThrProAlaLys 130
267 GCTACTCGCATAAAGTACACTGGCAAGACAGCTATCAGAGCTGACG 316
|||||:|||||:|||||:|||||:
131 G1yAlaLysAsnGlyLysAsn.AlalYsLysGluAspSerAspGluAsp 147
317 AGAAGAAGAAAGCAGACAGAGAGACAACAACAGCAGCTGAGAGCG 366
|||||:|||||:|||||:|||||:
147 LuAspAspAspAspAspAspLysAspSerAspGluAspGluGluAsp 163
164 GluLysAspGluPheGluProValYalLysGlyLys...GlnGlyLys 179
417 AGTA.....ATCTCAGTTGACGAGAGACCAAAACGAGAG 451
|||||:|||||:|||||:|||||:
179 sValAlaAlaAlaAlaProAlaSerGluAspGluAspGluGluAsp 196
452 ACATGTTGATCAAGAAACGAGCGGAGACAGACTGTGCAAGAGTGC 501
|||||:|||||:|||||:|||||:
196 LuGluGluGluGluGluGluAspGluGluGluGluAspSerGluGlu 212
502 AGGCG 507
|||:|||||:
213 GluAla 214

```

seq_name: SwissProt_39:GAG_MSVMO

seq_documentation_block:

ID GAG_MSVMO STANDARD; PRT; 538 AA.

AC P03334;

DT 21-JUL-1986 (Rel. 01, Created)

DT 21-JUL-1986 (Rel. 01, Last sequence update)

DT 20-AUG-2001 (Rel. 40, Last annotation update)

DE GAG POLYPROTEIN R65 [CONTAINS: CORE PROTEIN P15; INNER COAT PROTEIN P12; CORE SHELL PROTEIN P30; NUCLEOPROTEIN P10].

DE DE GAG.

OS Moloney murine sarcoma virus.

OC Viruses; Retroviral viruses; Retroviridae; Gammaretrovirus.

OX NCBI_TaxID=11809;

RN [1]

RP SEQUENCE FROM N.A. (PROVIRUS).

```

RX MEDLINE=62039559; PubMed=6170110;
RA Reddy E.P., Smith M.J., Aaronson S.A.;
RT "Complete nucleotide sequence and organization of the Moloney murine
RL sarcoma virus genome.";
RN Science 214:445-450(1981).
RP SEQUENCE FROM N.A. (CLONE 124 CIRCULAR).
RX MEDLINE=62115347; PubMed=6173134;
RA van Beveren C., van Straten F., Galleshaw J.A., Verma I.M.;
RT "Nucleotide sequence of the genome of a murine sarcoma virus.";
RN Cell 27:97-108(1981).
RP SEQUENCE OF 2-17, AND MYRISTOYLATION.
RX MEDLINE=63169654; PubMed=6340098;
RA Henderson L.E., Kruttsch H.C., Oroszlan S.;
RT "Myristyl amino-terminal acylation of murine retrovirus proteins: an
RT unusual post-translational protein modification.";
RL Proc. Natl. Acad. Sci. U.S.A. 80:339-343(1983).
CC -1- PTM: SPECIFIC ENZYMAITC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL: J02266; AAA46499.1; -.
DR EMBL: V01185; CAA24507.1; -.
DR PIR: A03932; FOMVM.
DR InterPro: IPR000840; Gag_MA.
DR InterPro: IPR002079; Gag_P12.
DR InterPro: IPR003036; Gag_P30.
DR InterPro: IPR001878; ZnF_CCHC.
DR Pfam: PF01140; gag_MA; 1.
DR Pfam: PF01141; gag_P12; 1.
DR Pfam: PF02093; Gag_P30; 1.
DR Pfam: PF00098; zf-CCHC; 1.
DR SMART: SM00343; ZnF_C2HC; 1.
KW Coat protein; Core protein; Nucleoprotein; Polypeptide; Myristate.
FT CHAIN 2 131 CORE PROTEIN P15.
FT CHAIN 132 215 INNER COAT PROTEIN P12.
FT CHAIN 216 478 CORE SHELL PROTEIN P30.
FT CHAIN 479 534 NUCLEOPROTEIN P10.
FT LIPID 2 2 MYRISTATE.
FT VARIANT 519 519 R -> K (IN CLONE 124).
SQ SEQUENCE 538 AA: 61209 MW: D78326F3B5701E56 CRC64;

```

alignment_scores:

Quality: 91.00 Length: 163
Ratio: 1.071 Gaps: 9
Percent Similarity: 52.147 Percent Identity: 26.380

alignment_block:
US-09-823-101-4 x GAG_MSVMO ..

Align seg 1/1 to: GAG_MSVMO from: 1 to: 538

```

84 AAAAAAGGAGAAACGAGGAGCATGAGTGTCCGGG.....CT 127
|||||:|||||:|||||:|||||:
353 LysValLysGlyIleThrGlnGlyProAsnGluSerProSerAlaPheLe 369
128 GGAACCTGTGTCACCGGCTACAA.....A 153
|||||:|||||:|||||:|||||:
369 uGluArgLeuLysGluAlaTyrArgArgTyrThrProTyrAspProGlu 386
154 CTCGAAAGAGAAACATTTGCAACGCGCTCCACACTACGCGGACAGACT 203
|||||:|||||:|||||:|||||:
386 sProGlyGlnGluThrAsnValSerMetSerPheIleTyrPheIleAla 402
204 .....AGAGAAACACATCTACTA..... 221

```

```

403 proaspllelglyarplysleugluarplyleugluaspleuargasnlysth 419
222 .....GAAATCACAACATATATATAGTACAGAG 252
419 rleuglyaspleuvalarglualegluargliepheasn...lysarvg 435
253 AAAGCCACATAGAGCTACTGCGATAAGTACAACGCGAAGACAGCT 302
435 luthrpfoglu.....Arggluarglu 443
303 ATCAGAGACTGACGAGAAAGAAACGACGACAGAGAACCAACGA 352
444 lleararggluarplysleugluarplyarplyrthrargluaspl 460
353 GCGAGCGTCAGCGGATGACGACGACGATACCGCAAGAAACGGGG 402
460 u.Glnlysgluys...gluarasparplyarplyhsalrglumetser 475
403 AAAGACAGATGAAGTAACTCATCTGAGTACGAAAGCAACCAAGAGA 452
476 ArgleuLeuAlaThrValValser.....Glygluargluaspar 489
453 CGATGTTGATCAAGAAACCGACGCGACAGACTGT 489
489 ggingluargluarplyarplyserglnleuaspys 501

```

seq_name: SwissProt_39:INCE_CHICK

seq_documentation_block:

ID INCE_CHICK STANDARD: PRT: 877 AA.

AC P53352;

DT 01-OCT-1996 (Rel. 34, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 30-MAY-2000 (Rel. 39, Last annotation update)

DE INNER CENTROMERE PROTEIN (INCENP).

OS Gallus gallus (Chicken).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Gallus.

OC NCBI_TaxID=9031;

RN [1]

RA MEDLINE=94012983; PubMed=8408220;

RA Mackay A.M., Eckley D.M., Chue C., Earnshaw W.C.;

RT "Molecular analysis of the INCENPs (inner centromere proteins):

RT separate domains are required for association with microtubules

RT during interphase and with the central spindle during anaphase.";

RL J. Cell Biol. 123:373-385(1993).

-1- FUNCTION: MAY ACT AS CYTOSKELETAL PROTEINS THAT ARE POTENTIALLY

CC CAPABLE OF ALTERING THE MORPHOLOGY OF THE CELLULAR MICROTUBULE

CC NETWORK DURING INTERPHASE.

-1- SUBUNIT: HOMODIMER OR HETERODIMER (PROBABLE).

CC -1- SUBCELLULAR LOCATION: NUCLEAR. IS RESTRICTED TO THE NUCLEUS IN

CC INTERPHASE, REMAINS TIGHTLY BOUND TO THE CHROMOSOMES UNTIL EARLY

CC METAPHASE, AND DURING LATE METAPHASE GETS CONCENTRATED IN LINEAR

CC ARRAYS THAT TRANSECT THE METAPHASE PLATE BETWEEN THE CHROMOSOMES.

CC AS ANAPHASE BEGINS TO MOVE TO THE SPINDLE MIDZONE WHERE IT IS

CC INTIMATELY ASSOCIATED WITH THE BUNDLED MICROTUBULES. LATER IN

CC ANAPHASE GETS CLOSELY ASSOCIATED WITH THE CELL CORTEX, AND BY

CC TELOPHASE IS CONCENTRATED AT EACH SIDE OF THE MIDBODY IN THE

CC INTERCELLULAR BRIDGE, WITH WHICH IT IS DISCARDED AFTER.

CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; INCENP I AND INCENP II (SHOWN

CC HERE); ARE PRODUCED BY ALTERNATIVE SPLICING.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

```

DR EMBL; Z25419; CAA80906.1; -.
KW Cell division; Microtubules; Coiled coil; Centromere; Mitosis;
KW Cell cycle; Nuclear protein; Alternative splicing.
FT DOMAIN 503 733 COILED COIL (POTENTIAL).
FT VARSPPLIC 716 753 MISSING (IN ISOFORM INCENP I).
FT VARIANT 257 257 A -> T.
FT VARIANT 471 471 E -> Q.
SQ SEQUENCE 877 AA; 100940 MM; AFA703149F55352 CRC64;

```

alignment_scores:

Quality:	Ratio:	Length:
91.00	0.948	179
Percent Similarity: 53.631	Percent Identity: 20.670	Gaps: 5

alignment_block:

US-09-823-101-4 x INCE_CHICK ..

Align seg 1/1 to: INCE_CHICK from: 1 to: 877

```

40 AAACGTAATCTATCAAGAGCAAAAAA.....AAACACC 89
523 ArglysglnlyvalgluargluyslysarplyarplylgluAla 539
90 CCGGCGCAACAGCGGACATACGAGTCGCCGGCTGGAACCTGTGT 139
539 sleuargargluargluargluarglyval..... 549
140 ACCCGGCTACAAAATCTCAAAAGAGCAACATTGTCAACGCGCTCACTA 189
550 .....LeuGlnAlaArg.....GluArgAlaGluGln 558
190 CCGGCGACACAGCTAGAGAAACACATCTACTAGAACTCAACATATAT 239
559 LeuGluGluGluArglysarplyarplyleuGluGluAlaLeuPheas 575
240 ACATAGTACAGAGAACGCCACATAGACTACTCGCA.....T 277
575 pgluylsThrGlnlysaAlaargluargluargluAlaGluGlnlysl 592
278 AAAGTCACTGCGCAAGACAGCTATCAGAGCTGACGACAGACAGAAC 327
592 ylsysarAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 608
328 GAGGAC..... 333
609 GlnasplgluAlaAlaarglyGlnlysaAlaLeuGluGluGluGlu 625
334 .ACAGAGACAAACCAACAGCGAGCGGTGACGCGATGAGAGCAGCGCA 382
625 uArgarplyhslysgluLeumetGlnlyslslysgluGluGluGluGln 642
383 GATACCCGAAAGAACGGGGAAGAACAGATG.....AAA 417
642 rglAlaarglyslleAlaGluargluargluAlaGluGluGluGluGln 658
418 GTATCTCACTTGCAGAGAACCAACCAACCGAGAGAGATCTTATCAAG 467
659 GlnleuAlaAlaGluargluargluGlnlysaAlaGlysgluGlnly 675
468 AAACGAGCGGACAGAGACTGTGACAGAGGTGGAGG 504
675 slslysgluGluargluargluAlaGluAlaGluysGlnarg 687

```

seq_name: SwissProt_39:GAG_MLVHO

seq_documentation_block:

ID GAG_MLVHO STANDARD: PRT: 540 AA.

AC P21435; Q80877; Q80878; Q80879; Q80880;

DT 01-MAY-1991 (Rel. 18, Created)

DT 01-MAY-1991 (Rel. 18, Last sequence update)

DT 20-AUG-2001 (Rel. 40, Last annotation update)

DE GAG POLYPROTEIN [CONTAINS: CORE PROTEIN P15; INNER COAT PROTEIN P12;
 DE CORE SHELL PROTEIN P30; NUCLEOPROTEIN P10].
 GN GAG.
 OC HOMUly murine leukemia virus (Mus hortulanus virus).
 OC Viruses. Retrovirdae; Retroviridae; Gammaretrovirus.
 ON NCBL_TaxID=11799;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=90051094; Pubmed=2554579;
 RX Voytek P., Kozak C.A.;
 RT "Nucleotide sequence and mode of transmission of the wild mouse
 ecotropic virus, HomuLV".
 RL Virology 173:58-67(1989).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M26528; AAA5464.1; -
 DR PIR: A32594; FOMVHL.
 DR InterPro: IPR000840; Gag_MA.
 DR InterPro: IPR002079; Gag_P12.
 DR InterPro: IPR003036; Gag_P30.
 DR InterPro: IPR001878; znf_CCHC.
 DR Pfam: PF01140; gag_MA.1.
 DR Pfam: PF01141; gag_P12.1.
 DR Pfam: PF02093; Gag_P30.1.
 DR Pfam: PF00098; znf_CCHC.1.
 DR SMART: SM00343; znf_CCHC.1.
 KM Core protein: Coat protein; Nucleoprotein; Polyprotein; Myristate.
 FT CHAIN 2 129 CORE PROTEIN P15.
 FT CHAIN 130 217 INNER COAT PROTEIN P12.
 FT CHAIN 218 480 CORE SHELL PROTEIN P30.
 FT CHAIN 481 540 NUCLEOPROTEIN P10.
 FT LIPID 2 2 MYRISTATE (BY SIMILARITY).
 SQ SEQUENCE 540 AA; 61382 MW; 5F25150086F9E95 CRC64;

alignment_scores:
 Quality: 90.50 Length: 165
 Ratio: 1.090 Gaps: 10
 Percent Similarity: 50.303 Percent Identity: 29.091

alignment_block:
 US-09-823-101-4 x GAG_MLVHO ..

Align seg 1/1 to: GAG_MLVHO from: 1 to: 540

```

84 AAACCCGGGGGAAACGAGCATGACGCGG.....CT 127
   |||:|||||:|||||:|||||:|||||:|||||:  ||
   |||:|||||:|||||:|||||:|||||:|||||:  ||
355 LysValValysGlyIleThrgInglProAsngInserProSerAlaPhele 371
   |||:|||||:|||||:|||||:|||||:|||||:  A 153
128 GGAACAGCTGGTACCCGGCTACAAA.....A 153
   |||:|||||:|||||:|||||:|||||:|||||:  A 153
371 uGluArgLeuLysGluAlaItyrArgItyrThrProTyrAspProGluA 388
154 CTCCAAGAGAGAAACATGTTCACGCGCTCCACTACCGCGACACAGCT 203
   |||:|||||:|||||:|||||:|||||:|||||:  ||
388 sPrroGlyGIngluThrAsnValSerMetSerPheIletrPInSerAla 404
204 .....AGAGAAACATCTACTA..... 221
405 ProAspIleGlyArgGlyLysLeuGluArgLeuGluAspLeuLysSerLysTh 421
222 .....GAAACATCAACATATATACATAGTACAGAG 252
421 rIeuGlyAspLeuValArgGluAlaGluArgIlePheAsn..LysArgG 437
253 AAACGACATAGAGCTACTGCGATTAAGTACACTGCGCAAGACACGCT 302

```

```

437 |||:|||||:|||||:|||||:|||||:|||||:  ArgGluGluArg 445
437 luThrProGluGlu.....  ArgGluGluArg 445
303 ATTCAGACTGACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 352
   |||:|||||:|||||:|||||:|||||:|||||:  |||
446 lIleArgArgGluThrGIngluLysGluGluArgArgAlaGluAsnGl 462
353 GCCACGCTGAGCGCGATGAGCAGACGCGAGATACCCGAAAGAGAGCGGG 402
   |||:|||||:|||||:|||||:|||||:|||||:  |||
462 u.GluArgGluLys...GluArgAspArgArgArgHisArgGluMetSer 477
403 AAAGACAGATGAAGTAATCTCAGTTCAGACGAGACCAAGAGAGAGAGA 452
   |||:|||||:|||||:|||||:|||||:|||||:  |||
478 LysLeuLeuAlaThrValValSer...GlyGluArgGluAspArgGInGl 493
453 CGATGTCATCAAGAAACGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 492
493 yGlyGluArgArgArgProGInLeuAspLysAspGInLysAla 507
seq_name: SwissProt_39:YL02_CAEEL
seq_documentation_block:
ID YL02_CAEEL STANDARD; PRT; 304 AA.
AC P34396;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-JUN-1994 (Rel. 28, Last annotation update)
DE HYPOTHETICAL 33.9 KDA PROTEIN F10E9.2 IN CHROMOSOME III.
GN F10E9.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
ON NCBL_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE=94150718; Pubmed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copesey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,
RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,
RA Johnston L., Jones M., Kershaw J., Kirsten J., Laister N.,
RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkhen R.,
RA Sims M., Smaldon N., Smith A., Smith M., Sonhammer E., Staden R.,
RA Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
RA Waterson R., Watson A., Weinstock L., Wilkinson-Sproat J.,
RA Wohlman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans."
RL Nature 368:32-38(1994).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  

CC or send an email to license@isb-sib.ch).  

CC -----
DR EMBL: L10986; AAA28013.1; -
DR PIR: S44801; S44801.
DR WormPep: F10E9.2; CE00146.
KW Hypothetical protein.
SQ SEQUENCE 304 AA; 33925 MW; 60999E5AE3402A2F CRC64;


alignment_scores:  

  Quality: 90.00 Length: 156  

  Ratio: 1.139 Gaps: 5  

  Percent Similarity: 50.641 Percent Identity: 23.077



alignment_block:


```

US-09-823-101-4 x YLU2_CAEL ..

Align seg 1/1 to: YLU2_CAEL from: 1 to: 304

```

34 GGGGAAAAAGTAAATCTATCAAGAGGAAAAA.....AAAAA 83
138 GilylslelrvglyserthrargalathrarglyslervalaInly 154
84 AAAACCCGGGGGAAACCGAGGACATAGCGGTCCGGCTGGAAC 133
154 sserlysserGlyArgThrArgHlsLysLysValProAspGlnArgL 171
134 TGTGTATC.....CCGGCTACAAAAATCCAAAGAGG 165
171 eutrpyrphleuTyLrLeuSerlyspValVal..... 182
166 AAMCATTTGTCACGCGCTCCACTACCGCGCACACAGCTAGAGAACAT 215
183 .....ValSerLeuProThrAspAlaThrTyrrHlsAspLe 194
216 CTACTAGAAACTCAACACATATATACATAGTAGAGAAACGCACATAG 265
194 uMetAlaArgLeuTyThrLeuLeuGlyValSerGlyLysThrHlsMetI 211
266 AGCTACTCGCATAAAGTACAACTGCGCAAGACACGTATCAGAGACTGAC 315
211 lethrleu.....ArgThrGlnLysSerSerSerGlnGly 223
316 GAGAGAAAGAAACGACGACAGAGAGAACCAACGACGCGCTGAGCG 365
224 GlnGlyGlnhserGlyArgGlyAsp...ProAlaProLeuGlnValG1 239
366 CGATGACGAGGACGACAGATACCCGAAAGACGGGAAAGACAGATGA 415
239 yAspSerlyThrArgLeuSerAspLeuLysIleThrArgHlsLysMetI 256
416 AAGTAATCTCAGTTGACGAGAACCAAAAGCGAGAGACGATGTTGATCA 465
256 euteleuAspValAsp...AspGlnLysLysAspGlySerGlyGlnGlu 271
466 AGAAACCGCGCGAGACA 483
272 LysLysGlnLysLysSer 277

seq_name: swissprot_39:ATRX_HUMAN
seq_documentation_block:
ID ATRX_HUMAN STANDARD: PRT: 2375 AA.
AC P46100; P51068; Q15886;
DT 01-NOV-1995 (Rel. 32, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE TRANSCRIPTIONAL REGULATOR ATRX (X-LINKED HELICASE II) (X-LINKED
DE NUCLEAR PROTEIN) (XNP).
GN ATRX OR RAD54L OR XH2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97123494; PubMed=89968741;
RA Picketts D.J., Higgs D.R., Bachoo S., Blake D.J., Quarrell O.W.,
RA Gibbons R.J.;
RT "ATRX encodes a novel member of the SNF2 family of proteins: mutations
RT point to a common mechanism underlying the ATR-X syndrome.";
RL Hum. Mol. Genet. 5:1899-1907(1996).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=97385582; PubMed=9244431;
RA Villard L., Lossi A.M., Cardoso C., Proud V., Chironi P.,
RA Colleaux U., Schwartz C., Fontes M.;
RT "Determination of the genomic structure of the XNP/ATRX gene encoding

```

```

RT a potential zinc finger helicase.";
RL Genomics 43:149-155(1997).
RN [3]
RP SEQUENCE OF 743 FROM N.A.
RX MEDLINE=95179111; PubMed=7874112;
RA Stayton C.U., Dabovic B., Gullisano M., Geetz J., Broccoli V.,
RA Giovannazzi S., Bossolasco M., Monaco L., Rastan S., Boncinelli E.,
RA Bianchi M.E., Consalez G.G.;
RT "Cloning and characterization of a new human Xq13 gene, encoding a
RT putative helicase.";
RL Hum. Mol. Genet. 3:1957-1964(1994).
RN [4]
RP PRELIMINARY PARTIAL SEQUENCE FROM N.A.
RX MEDLINE=94214473; PubMed=8162050;
RA Geetz J., Pollard H., Consalez G., Villard L., Stayton C.,
RA Millaesau P., Khrestchatsky M., Fontes M.;
RT "Cloning and expression of the murine homologue of a putative human
RT X-linked nuclear protein gene closely linked to FCK1 in Xq13.3.";
RL Hum. Mol. Genet. 3:39-44(1994).
RN [5]
RP SEQUENCE OF 2284-2375 FROM N.A., AND VARIANTS ATR-X.
RX MEDLINE=95211835; PubMed=7697714;
RA Gibbons R.J., Picketts D.J., Villard L., Higgs D.R.;
RT "Mutations in a putative global transcriptional regulator cause X-
RT linked mental retardation with alpha-thalassemia (ATR-X syndrome).";
RL Cell 80:837-845(1995).
RN [6]
RP VARIANTS ATR-X.
RX MEDLINE=97467722; PubMed=9326931;
RA Gibbons R.J., Bachoo S., Picketts D.J., Afimos S., Azenbauer B.,
RA Bergoffen J., Berry S.A., Dahl N., Fryer A., Keppler K., Kurosawa K.,
RA Levin M.L., Masuno M., Neri G., Pierpont M.E., Slaney S.F.,
RA Higgs D.R.;
RT "Mutations in transcriptional regulator ATRX establish the functional
RT significance of a PHD-like domain.";
RL Nat. Genet. 17:146-148(1997).
RN [7]
RP VARIANT JM GLN-2014.
RX MEDLINE=96224392; PubMed=8630485;
RA Villard L., Geetz J., Mattei J.-F., Fontes M., Saugier-Verber P.,
RA Munnich A., Lyonnet S.;
RT "XNP mutation in a large family with Juberger-Marsidi syndrome.";
RL Nat. Genet. 12:359-360(1996).
RN [8]
RP VARIANT ATR-X LEU-129.
RX MEDLINE=20123062; PubMed=10660327;
RA Fichera M., Romano C., Castiglita L., Falla P., Ruberto C., Amata S.,
RA Greco D., Cardoso C., Fontes M., Ragusa A.;
RT "New mutations in XNP/ATR-X gene: a further contribution to
RT genotype/phenotype relationship in ATR/X syndrome.";
RL Hum. Mutat. 12:214-214(1998).
CC -1- FUNCTION: COULD BE A GLOBAL TRANSCRIPTIONAL REGULATOR. MODIFIES
CC GENE EXPRESSION BY AFFECTING CHROMATIN. MAY BE INVOLVED IN
CC BRAIN DEVELOPMENT AND FACIAL MORPHOGENESIS.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- DISEASE: DEFECTS IN ATRX ARE THE CAUSE OF X-LINKED ALPHA-
CC THALASSEMIA/MENTAL RETARDATION SYNDROME (ALSO KNOWN AS ATR-X
CC SYNDROME). ATR-X IS AN X-LINKED DISORDER COMPISING SEVERE
CC PSYCHOMOTOR RETARDATION, CHARACTERISTIC FACIAL FEATURES, GENITAL
CC ABNORMALITIES, AND ALPHA-THALASSEMIA.
CC -1- DISEASE: DEFECTS IN ATRX ARE THE CAUSE OF JUBERGER-MARSIDI SYNDROME
CC (JM). JM IS A RARE X-LINKED RECESSIVE DISEASE CHARACTERIZED BY
CC SEVERE MENTAL RETARDATION, GROWTH FAILURE, SENSORINEURAL DEAFNESS,
CC MICROGENITALISM AND EARLY DEATH.
CC -1- SIMILARITY: BELONGS TO THE SNF2/RAD54 HELICASE FAMILY.
CC -1- SIMILARITY: CONTAINS 1 PHD-TYPE ZINC FINGER.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/

```


1364 ArgLysIleLeuLysAspAspLysLeuArgThrGluThrGlnAsnAlaLe 1380
 459 TGATCAAGAAGAACCGCGACGACAGACAGTGTGCAGAGTGCGAGG 504
 : ::::::::::::::: ||||| |||
 1380 uLysGluGluGluArgArgLysArgLysAlaGluArgGluArg 1395

seq_name: SwissProt_39:MST2_DROHY

seq_documentation_block:

ID MST2_DROHY STANDARD; PRT; 1391 AA.

AC 008696;

DT 01-FEB-1995 (Rel. 31, Created)

DT 01-FEB-1995 (Rel. 31, Last sequence update)

DT 20-AUG-2001 (Rel. 40, Last annotation update)

DE AXONEME-ASSOCIATED PROTEIN MST101(2).

GN MST101(2).

OS Drosophila hydei (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OX Ephydroidea; Drosophilidae; Drosophila.

OX NCBI_TaxID=7224;

RN [1]

RP SEQUENCE FROM N.A., AND CHARACTERIZATION.

RX MEDLINE=95045538; PubMed=7957199;

RA Neesen J., Padmanabhan S., Buenemann H.;

RT "Tandemly arranged repeats of a novel highly charged 16-amino-acid

RT motif representing the major component of the sperm-tail-specific

RT axoneme-associated protein family Dms101 form extended

RT alpha-helical rods within the extremely elongated spermatozoa of

RT Drosophila hydei."

RL Eur. J. Biochem. 225:1089-1095(1994).

CC -1- FUNCTION: POSSIBLE STRUCTURAL ROLE IN THE SPERM TAIL.

CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.

CC -1- TISSUE SPECIFICITY: TESTIS. PRIMARY SPERMATOCTYES AND EARLY

CC SPERMATIDS.

CC -1- DOMAIN: THE PREDOMINANT STRUCTURE IS ALPHA-HELICAL.

CC -1- POLYMORPHISM: LENGTH POLYMORPHISMS EXIST BETWEEN DIFFERENT

CC STRAINS, MOST LIKELY CAUSED BY LENGTH VARIATIONS WITHIN THE TANDEM

CC REPEATS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation-

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@sib-sib.ch).

CC -----

DR EMBL: X73481; CAA51876.1; -.

DR PIR: S34154; S34154.

DR HSSP: P01032; 1C5A.

DR FlyBase: FBgn0020733; Dhydvmst101(2).

KW Sperm: Repeat: Multigene family; Polymorphism.

FT DOMAIN 332 1268 59 X 16 AA APPROXIMATE TANDEM REPEATS OF

FT [KR]-K-X-C-X-X-A-K-X-X-K-X-X-X-E.

SO SEQUENCE 1391 AA; 159000 MW; 1B2A368F30F4878 CRC64;

alignment_scores:

Quality: 88.50 Length: 200

Ratio: 0.903 Gaps: 6

Percent Similarity: 49.000 Percent Identity: 19.500

alignment_block:

US-09-823-101-4 x MST2_DROHY ..

Align seg 1/1 to: MST2_DROHY from: 1 to: 1391

1 CAGGAATTACTTAACAGATGCTATATCATCATCGAGAAAAACGTAATTC 50
 ::::::::::::::: ||||| ::
 720 GIUGluLeuAlaLysLysIleLysAlaAlaGluLysLysCysLys 736
 51 TATCAAGAGAGAAAAA.....AAAAAAAAAAAAAAAAACCG 91

736 sLysLeuAlaLysLysLysLysAlaGluLysAsnLysLeuLysLysG 753
 92 GGGGGAACACGAGGACATACGAGTGTCCCGGCTGGAAACATGTGTATC 141
 ||::||| |||
 753 LysnLysLysGlyLysLysAla..... 760
 142 CCGGCTCAAAACTCCAAAGAGAAACATTGTCAACGCGCTCCACTACC 191
 ||::::::::||| ||:::
 761LeuLysGluLysLysCysArgLysLysLysLysLys 773
 192 GCCGCACACAGCTAGAGAACACATCTACTAGAAATCTCAACATATATAC 241
 ::::: ::|||
 773 sLysAlaAlaGluLysLysLysCysLysGluAlaAlaLysLysGluLysG 790
 242 ATAGTACAGAGAACCCACATAGAGCTACTCCATTAATTAACACTCGC 291
 ::::::::::|||::: |||
 790 LuAlaAlaGluLysLysLysCysGlu.....LysThrAla 801
 292 CAAGACACGATATCAGAGCTGACGAGAGAGAAAGACGACAGAGAGA 341
 ::::: :::::
 802 LysLysArgLysGluGluAlaGluLysLysCysGluLysThrAlaLys 818
 342 CAAA.....CCAACGAGCCAGC 358
 ||| :::::
 818 sLysArgLysGluAlaAlaGluLysLysCysGluLysAlaAlaLysL 835
 359 GTAGCGCGATGAGCGACGACGATACCCGAAAGAACGGGGAAGA 408
 ::::: :::::
 835 ysArgLysGluGluAlaGluLysLysCysGluLysThrAlaLysLys 851
 409 CAGATGAAAGTAATCTCAGTTGACGAAGCAAGAGAGAGACATGT 458
 ::::: :::::
 852 ArgLysGluThrAla.....GluLysLysCysGluLysAlaLys 865
 459 TGATCAAGAAGAACGAG..... 474
 : ::|||::|||
 865 aLysLysArgLysGluAlaAlaGluLysLysCysGluLysAlaAlaLys 882
 475CGGAGACAGACCTGTCCAGAAAGTGGAGG 504
 ::::: |||||:::
 882 ysLysArgLysGluAlaAlaGluLysLysCysAlaGluAlaLys 898

THIS PAGE BLANK (USPTO)

OM of: US-09-823-101-4 to: SPTRMBL_17.* out_format : pfs
Date: Jan 17, 2002 4:07 PM

About: Results were produced by the Gencore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

```
-MODEL=frame+u2p.model -DEV=xlp  
-O=/cgn2.1/USPRO_pool/US09823101/runat_17012002.155138.22043/app-query.fasta_1.2239  
-DB=SPTRMBL_17 -DBMT=fastan -SUFFIX=n2p rspt -GAPOP=12.000  
-GAPEXT=4.000 -MINMATCH=0.100 -LOOCL=0.000 -LOOEXT=0.000  
-GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500  
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500  
-DELPO=6.000 -DELEXT=7.000 -START=1 -MATRIX=blomsu62  
-TRANS=human40.cdi -LIST=45 -DOCLIGN=200 -THR_SCORE=pcr  
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs  
-NOR-ext -MINLEN=0 -MAXLEN=200000000  
-USER=US09823101 -ECGN1_1.384 -NCP0=6 -ICPU=3 -LONGLOG -NO_XLPXY  
-WAIT -THREADS=1
```

Search information block:

Query: US-09-823-101-4
Database: SPTRMBL_17.*
Database sequences: 473505
Database length: 146272329
Search time (sec): 224.230000

Sequence	Strd Orig	ZScore	EScore	Len	Documentation
sp_plant:09FR60	109.50	173.53	0.0144	1501	09fr60 oryza sativa (rice). md
sp_rident:09GXK6	109.50	173.02	0.0142	1620	09gxk6 mus musculus (mouse). m
sp_rident:09DBR7	104.00	166.66	0.0519	1004	09db7 mus musculus (mouse). m
sp_invertebrate:09NAU5	103.00	169.00	0.0705	547	09na5 caenorhabditis elegans
sp_invertebrate:09VB01	103.00	164.88	0.0648	1010	09vb1 drosophila melanogast
sp_invertebrate:024898	102.00	167.80	0.0891	505	024898 ensis minor (razor sh
sp_plant:09LW95	101.50	166.82	0.0994	513	09lw95 nicotiana tabacum (comm
sp_plant:049677	100.50	159.03	0.1099	1260	049677 arabidopsis thaliana (m
sp_invertebrate:09W3D1	100.00	158.59	0.1240	1183	09w3d1 drosophila melanogast
sp_bacteria:087306	99.00	164.04	0.1795	406	087306 borrelia burgdorferi (f
sp_plant:09LH98	99.00	153.04	0.1435	2081	09lh98 arabidopsis thaliana (m
sp_invertebrate:023319	98.50	163.46	0.2019	389	023319 caenorhabditis elegans
sp_virus:055708	98.50	162.39	0.1975	456	055708 chilo iridescent virus
sp_invertebrate:017840	98.50	161.91	0.1956	490	017840 caenorhabditis elegans
sp_plant:09S174	98.00	157.44	0.2033	836	09s174 arabidopsis thaliana (m
sp_invertebrate:023330	97.50	160.12	0.2443	493	023330 caenorhabditis elegans
sp_invertebrate:0902M6	97.50	155.64	0.2230	960	0902m6 caenorhabditis elegans
sp_invertebrate:002061	97.50	153.30	0.2126	1359	002061 caenorhabditis elegans
sp_invertebrate:09U7E0	97.50	153.30	0.2126	1359	09u7e0 caenorhabditis elegans
sp_invertebrate:09M4J3	97.00	157.60	0.2642	630	09m4j3 drosophila melanogast
sp_fungi:09HFC4	96.50	161.04	0.3225	332	09hfc4 neurospora crassa. cons
sp_invertebrate:076719	96.50	160.98	0.3221	335	076719 caenorhabditis elegans
sp_rident:070487	96.50	160.88	0.3214	340	070487 rattus norvegicus (rat)
sp_invertebrate:09B2I7	96.50	158.52	0.3064	483	09b2i7 homo sapiens (human). u
sp_rident:092330	96.50	150.37	0.2595	1622	092330 rattus rattus (black r
sp_invertebrate:017909	96.00	159.17	0.3534	385	017909 caenorhabditis elegans
sp_invertebrate:09N1C1	96.00	158.11	0.3458	451	09n1c1 caenorhabditis elegans
sp_human:09H1J0	96.00	157.83	0.3438	470	09h1j0 homo sapiens (human). hu
sp_plant:09FWA0	96.00	156.57	0.3351	567	09fwa0 arabidopsis thaliana (m
sp_plant:09M8A0	96.00	156.36	0.3337	585	09m8a0 arabidopsis thaliana (m
sp_invertebrate:09V2G2	96.00	155.91	0.3307	625	09v2g2 drosophila melanogast
sp_human:09N8T7	96.00	152.57	0.3089	1027	09n8t7 homo sapiens (human). h
sp_human:09S8I9	96.00	151.72	0.3036	1165	09s8i9 homo sapiens (human). h
sp_human:07S172	96.00	151.67	0.3033	1175	07s172 homo sapiens (human). h
sp_rident:097820	96.00	151.34	0.3013	1233	097820 mus musculus (mouse). m
sp_human:09S033	96.00	151.21	0.3005	1257	09s033 homo sapiens (human). h
sp_invertebrate:09V952	95.50	154.03	0.3622	726	09v952 drosophila melanogast
sp_invertebrate:061105	95.00	154.45	0.4158	600	061105 trypanosoma cruzi. try
sp_plant:092U69	95.00	153.42	0.4072	699	092u69 arabidopsis thaliana (m
sp_plant:092W13	95.00	153.43	0.3990	810	092w13 cucurbita maxima (pumpk

sp_invertebrate:061164 + 95.00 147.43 0.3605 1701 061164 plasmodium yoelii
sp_human:09H231 + 95.00 146.14 0.3511 2062 09h231 homo sapiens (human)
sp_invertebrate:091570 + 94.50 157.98 0.5085 312 091570 caenorhabditis eleg
sp_invertebrate:019973 + 94.50 156.76 0.4960 374 019973 caenorhabditis eleg
sp_rident:09JUK7 + 94.50 154.88 0.4774 494 09jk17 rattus norvegicus (ra

seq_name: sp_plant:09FR60

seq_documentation_block:
ID 09FR60 PRELIMINARY; PRT: 1501 AA.

AC 09FR60;
DT 01-MAR-2001 (TEMBLrel, 16, Created)
DT 01-MAR-2001 (TEMBLrel, 16, Last sequence update)
DE 01-JUN-2001 (TEMBLrel, 17, Last annotation update)
DE MODIFICATION METHYLASE (EC 2.1.1.73) (CYTOSINE-SPECIFIC
DE METHYLTRANSFERASE).
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. INDICA;
RA wiriyawitkorn N., Naraengajavana J.;
RT "Isolation and characterization of the putative cDNA encoding cytosine
RT DNA methyltransferase from rice (Oryza sativa).";
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
CC -! CATALYTIC ACTIVITY: S-ADENOSYL-L-METHIONINE + DNA CYTOSINE = S-
CC ADENOSYL-L-HOMOCYSTEINE + DNA 5-METHYLCYTOSINE.
CC -! SIMILARITY: TO C-5 CYTOSINE-SPECIFIC DNA METHYLASE FAMILY.
DR EMBL: AF155874; AAC43376.1; -
DR InterPro: IPR001025; BAH.
DR InterPro: IPR001525; C5_DNA_meth.
DR InterPro: IPR002857; ZnF-CXXC.
DR Pfam: PF01426; BAH; 2.
DR Pfam: PF00145; DNA_methylase; 3.
DR Pfam: PF02008; zf-CXXC; 1.
DR PRINTS: PR00105; C5METTFRASE.
DR SMART: SM00439; BAH; 2.
DR PROSITE: PS00094; C5_MTASE_1; 1.
KW Methyltransferase; Cytosine.
KW Methylintransferase; Transferase.
SQ SPOUNCE 1501 AA; 169588 MW; BDCFSIDC410BA3BD CRC64;

alignment_scores:
Quality: 109.50 Length: 148
Ratio: 1.335 Gaps: 7
Percent Similarity: 55.405 Percent Identity: 27.703

alignment_block:
US-09-823-101-4 x 09FR60 ..

Align seg 1/1 to: 09FR60 from: 1 to: 1501

```
54 CAAGAGGAAAAA.....AAAAA.....AAACCGGGGGAACACG 103  
:::||||: ::::| |||||::: ::::| |||||:  
74 GAGUUAAGAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAGG 90  
::: ::::| |||||::: ::::| |||||:  
104 GAGGACATGCGAGGAGTCTCCGGGCTGGAACGTCGTACCGCTAC 149  
::: ::::| |||||::: ::::| |||||:  
150 .....AAAGCTCAAAAGGAAAGGAAAGGAAAGGAAAGGAAAGG 191  
::: ||| |||||::: ::::| |||||:  
107 euglypPcGcGcGcGcGcGcGcGcGcGcGcGcGcGcGcGcGcG 123  
192 GCGACACAGCTAGAGAAACATCTACTAGAAACCTCAACATATATAC 241  
||||| ||||| ||| :::  
124 ArgHsHtr...ArgGluLeuSerLeuArgGlySer..... 135  
:::|||| ||| ||| ::::| |||
```

```

136 .....LysgluaspProaspArglualAArgProglu.Thr..... 147
292 CAAGAAGACCTATCAGAGCTGAGCAGAGAGAAAGCAGCAGAGAGAGA 341
148 .....HisleuaspgluaspgluaspGlyLysLysaspLysArgSerSe 162
342 CAAGCAAGCAGCAGCGCTGAGCGCGATGAGCAGAGCAGATATCCCGA 391
162 ArgProArgSerGln...ProArgaspProAlaAlaLysArgArgProL 178
392 AAGAAACGGGGGAAAGACAGATGAACTATCTCAGTTGACGAAGACCAA 441
178 ysgluAlaGluprogluGln.....ValAlaProgluThrProglu 191
442 AAGCGAGAGACGATGTTGATCAAGAAAGCAGCGCAGAGACA 483
192 AspArgaspLupLaspgluArggluGluLysArgArgLysThr 205
seq_name: sp_rodent:Q90XX6

```

```

seq_documentation_block:
ID Q90XX6 PRELIMINARY: PRT: 1620 AA.
AC Q90XX6:
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DE MODIFICATION METHYLASE (EC 2.1.1.73) (CYTOSINE-SPECIFIC
DE METHYLTRANSFERASE).
GN DNMT1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6;
RA Margot J.B., Aguirre-Arteta A.M., Di Giacomo V.B., Pradhan S.,
RA Roberts R.J., Cardoso M.C., Leonhardt H.;
RA "The genomic organization of the mouse DNA methyltransferase.";
RA Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: S-ADENOSYL-L-METHIONINE + DNA CYTOSINE = S-
CC ADENOSYL-L-HOMOCYSTEINE + DNA 5-METHYLCYTOSINE.
CC -1- SIMILARITY: TO C-5 CYTOSINE-SPECIFIC DNA METHYLASE FAMILY.
DR EMBL: AF162282; AAF19352.1; -
DR MGD: MGI:94912; Dnmt1.
DR InterPro: IPR001025; BAH.
DR InterPro: IPR001525; C5_DNA_meth.
DR InterPro: IPR002857; Znf-CXXC.
DR Pfam: PF01426; BAH; 2.
DR Pfam: PF00145; DNA_methylase; 3.
DR PRINTS: PF02008; zf-CXXC; 1.
DR SMART: SM00439; BAH; 2.
DR PROSITE: PS00094; C5_MTASE_1; 1.
DR PROSITE: PS00095; C5_MTASE_2; 1.
DR Mehlhyltransferase; Transferase.
KW SEQUENCE 1620 AA; 183187 MW; 4F9A98CEAF09F037 CRC64;

```

```

alignment_scores:
Quality: 109.50 Length: 148
Ratio: 1.335 Gaps: 7
Percent Similarity: 55.405 Percent Identity: 27.703

```

alignment_block:

US-09-823-101-4 x Q90XX6 ..

Align seg 1/1 to: Q90XX6 from: 1 to: 1620

```

54 CAAGAGCAAAAAAAAAAAAAAAAAAAAAACCCGGGGGGAACCGAG 103
193 GluGluArgaspLaspLysArgArgValAlaValAspThrGluSerG 209

```

```

104 GGGACATACGAGAGTGTCCCGGCTGGAAGTGTGATCCCGGCTAC.... 149
209 yAlaAlaAlaValAlaGluLysLeuGluGluValThrAlaGlyThrGlnL 226
150 .....AAACTCCAAAGAGAAACATTTGACACGCGCTCCACTACC 191
226 euGluprogluGluProGlyGluGluGlnGluLaspaspAsnArgSerLeuArg 242
192 GCGCACACAGCTGAGAGAACACATCTACTAGAAACTCACAACTATTATAC 241
243 ArgHisThr...ArgGluLeuSerLeuArgArgLysSer..... 254
242 ATAGTACAGAGAAACGCCATAGAGCTACTCCATTAAGTACAACTGCG 291
255 .....LysgluaspProaspArglualAArgProglu.Thr..... 266
292 CAAGAAGACCTATCAGAGCTGAGCAGAGAGAAAGCAGCAGAGAGAGA 341
267 .....HisleuaspgluaspgluaspGlyLysLysaspLysArgSerSe 281
342 CAAGCAAGCAGCAGCGCTGAGCGCGATGAGCAGAGCAGATATCCCGA 391
281 ArgProArgSerGln...ProArgaspProAlaAlaLysArgArgProL 297
392 AAGAAACGGGGGAAAGACAGATGAACTATCTCAGTTGACGAAGACCAA 441
297 ysgluAlaGluprogluGln.....ValAlaProgluThrProglu 310
442 AAGCGAGAGACGATGTTGATCAAGAAAGCAGCGCAGAGACA 483
311 AspArgaspLupLaspgluArggluGluLysArgArgLysThr 324
seq_name: sp_rodent:Q9DBR7

```

```

seq_documentation_block:
ID Q9DBR7 PRELIMINARY: PRT: 1004 AA.
AC Q9DBR7:
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE 1200015F06RIK.
GN 1200015F06RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=LUNG;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K.I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batilov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schiml L.M., Staibli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barish G.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Rint B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.F.,
RA Sasaki H., Toyooka K., Wang K.H., Weitz C., Wittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohlsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL: AK004785; BAB23563.1; -
DR MGD: MGI:1918986; 1200015F06RIK.
DR InterPro: IPR002110; ANK.

```

DR Pfam; PF00023; ank; 6.
DR SMART; SM00248; ANK; 6.
DR PROSITE; PS00088; ANK_REPEAT; 4.
DR PROSITE; PS0297; ANK_REPEAT_REGION; 1.
SO SEQUENCE 1004 AA, 111808 MW, 1CF711811B85BF1 CRC64;

alignment_scores:
Quality: 104.00 Length: 171
Ratio: 1.118 Gaps: 6
Percent Similarity: 54.386 Percent Identity: 24.561

alignment_block:

US-09-823-101-4 x Q9DBR7

Align seg 1/1 to: Q9DBR7 from: 1 to: 1004

```
33 AGGCAAAAACGTAATCTATCAGAGAGAAAAAAGAAAAA 82
|||||:|||||:|||||:|||||:|||||:|||||:
718 ArgGluGluGluAsnGluGluGluGluGluGluGluGlu 734
83 AAAACCCGGGGGGAACCGAGGAGATAGCGAGTCCCGGCTG 132
|||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
734 pLysGluGluGluGluGluGluGluGluGluGluGluGlu 751
133 CTGTGTACCGCGCTACAAACTCCAAAGAGAGAAACA 170
|||||:|||||:|||||:|||||:|||||:|||||:
751 LuTyTyGluGluGluSerArgThyThyArgGluThyThy 767
171 .....TTGTCAACGCGCTCCACTACCGCACACAGCTAGA 214
|||||:|||||:|||||:|||||:|||||:|||||:
768 ArgProValSerThySerSerSerSerAlaProSerSer 784
215 TCTACTAAACCTCACACATATATACATAGACAGAAACCA 264
|||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
784 ThrLeuGluSer..ThrLeuTyAlaSerSerGluLeuAsn 800
265 GAGTACTCGCATAAAGTAACTGCGCAAGACAGCTATCAGA 314
|||||:|||||:|||||:|||||:|||||:|||||:
801 SerLeuValGlyIleThrSerAlaTySerArgGlyLeuAl 817
315 CGAGAGAAAGAAACGACAGAGAGACAAACCAAGCCAGCT 364
|||||:|||||:|||||:|||||:|||||:|||||:
817 ngLuArg.....GluGlyGluGluGluGluGluGluGlu 830
365 GCGATGAG.....GCAAGCGCAGATACCCGAA 393
|||||:|||||:|||||:|||||:|||||:|||||:
830 LysIAspLysSerGluProLysSerIleArgGluArgArg 846
394 GAAACGGGGAAGAGACAGATGAAAGTAACTTCAGTT 429
|||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
847 Glu.....LysArgArgSerThrGlyValSerPheTrp 861
430 .....GACGAGACCAAAAGCGAGAGACGATTTGATCA 472
|||||:|||||:|||||:|||||:|||||:|||||:
861 rAspGluAsnGluGluGluArgGluGluSerPheGluAsp 878
473 AGCGGAGACA 483
|||||:|||||:|||||:|||||:|||||:|||||:
878 yArgGluThr 881

seq_name: sp_invertebrate:Q9N4U5
seq_documentation_block:
ID Q9N4U5 PRELIMINARY; PRT: 547 AA.
AC Q9N4U5;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DE 01-OCT-2000 (Tremblrel. 15, Last annotation update)
DE y49F6B.4 PROTEIN.
GN y49F6B.4
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;
```

OC Rhabditidae; Pelodidae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
RT Investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).

alignment_scores:
Quality: 103.00 Length: 143
Ratio: 1.321 Gaps: 5
Percent Similarity: 54.545 Percent Identity: 26.573

alignment_block:

US-09-823-101-4 x Q9N4U5

Align seg 1/1 to: Q9N4U5 from: 1 to: 547

```
60 GAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAG 109
|||||:|||||:|||||:|||||:|||||:|||||:
289 GluProLysLysLysLysLysGluGluPheArgValGlu 304
110 TAGCAGTGTCCCGGCTGGAACCTGTGCTACCGGCTACA 159
|||||:|||||:|||||:|||||:|||||:|||||:
304 rArgAspAlaProGlyAlaAlaAlaAlaProGlyAlaLys 318
160 AAGAGAAACATTTGTCAACGCGCTCCACTACCGCACAC 209
|||||:|||||:|||||:|||||:|||||:|||||:
319 .....GlyAsp 320
210 ACACATCTACTAGAAACCTCACACATATATACATAGTAC 259
|||||:|||||:|||||:|||||:|||||:|||||:
321 Ser..IleTyAspAspLeuAspAspTyValProSerArg 337
260 ACATAGAGTACTCGCATTAAGTACACCTGCGCAAGACAG 309
|||||:|||||:|||||:|||||:|||||:|||||:
337 sp.....SerArgAspAlaGlyArgArgGlySerArg 347
310 ACTGACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 359
|||||:|||||:|||||:|||||:|||||:|||||:
348 ArgAsp.....ArgSerArgAspArgSerArgAspArg 362
360 TGAGCGCGATGAGGACAGCCGACATACCCGAAAGAAACG 409
|||||:|||||:|||||:|||||:|||||:|||||:
362 gAspArgAspAsnArgAspArgTyPheGluGluSerAla 379
410 AGATGAAGTATCTCAGTTGACGAGAGAGAGAGAGAGAG 459
|||||:|||||:|||||:|||||:|||||:|||||:
379 rGluGluGluGluGluGluGluGluGluGluGluGluGlu 395
460 GATCAAGAAACCGAGGAGAGAGAGAGAGAGAGAGAGAG 486
|||||:|||||:|||||:|||||:|||||:|||||:
396 GluGluGluGluGluGluGluGluGluGluGluGluGlu 404

seq_name: sp_invertebrate:Q9VB01
seq_documentation_block:
ID Q9VB01 PRELIMINARY; PRT: 1010 AA.
AC Q9VB01;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DE 01-JUN-2001 (Tremblrel. 17, Last annotation update)
```


DR SMART: SM00526; H15; 1.
KW NUCLEAR PROTEIN.
SQ SEQUENCE 505 AA; 57911 MW; 21DBFDEB8533934B CRC64;

alignment_scores:
Quality: 102.00 Length: 159
Ratio: 0.962 Gaps: 5
Percent Similarity: 66.667 Percent Identity: 25.157

alignment_block:

US-09-823-101-4 x Q24898 ..

Align seg 1/1 to: Q24898 from: 1 to: 505

```
28 TCATCAGGCAAAACGTAATTCATACAGAGCAAAAAA 77
|||||  |||||||  |||||||  |||||||  |||||||
131 SerLaserLysArgSerLys.SerArgLysArgSerLaserArg 147
78 AAAAAAAACCCGGGGAACCAAGGAGCATAGCGAGTCCCGGCT 127
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
147 snlySerAsnAsnAsnThrAsnAsnSerAlaLysArgSerArgSer 163
128 GGAACAGTGTGTCACCGCTACAAACATCCAAAGAGCAACATTGTCAA 177
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||
164 ArgLysArgSerAlaSerLysArgSer...ArgSerArgLysArgSe 179
178 CGCGCTCCACATACCGGCACACACACTAGAGAAACACATCTACTAGAACT 227
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||
179 rLaserLysArgSerHisSerArgLysArgSerAlaSerLysLys 196
228 CACAACATATATACATAGTACAGAGAAACGCACATAGACTACCGCAT 277
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
196 rGSerHisSerArg....LysArgSerAlaSerLysArgSerHis 210
278 AAAGTACAACTGCGCAAAAGACACGATACAGACTGACAGAGAGAAAC 327
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||
211 SerArgLysArgSerAlaSerLysArgSerLysSerArgLysArgSe 227
328 GACACAGAGAGAGAACCAACCAAGCAGCGCTAGCGCGCATGAGCGCA 377
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
227 rAlaSerLysArgSerLysArgSerArgLysArgSerAlaSerLys 244
378 GCGCAGATACCGCA.....AAGAAACGGGAAAGACAGCA 412
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||
244 rGSerLysSerArgLysArgSerAlaSerLysArgSerLysArgSer 260
413 TGAAGTAACTCAGTTGACGAGAACCAAAAGCGAAGACAGATGTTGAT 462
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||
261 LysArgSerAlaSerLysArgSerLysArgSerLysArgSerAla 277
463 CAAGAAACCGCGCGAGACAGCA 485
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
277 erLysLysArgSerLysSerArg 284

seq_name: sp_plant:Q9LW95

seq_documentation_block:
ID Q9LW95 PRELIMINARY; PRT; 513 AA.
AC Q9LW95;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE KED.
GN KED.
OS Nicotiana tabacum (Common tobacco).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; eunasterids I; Solanales; Solanaceae; Nicotiana.
OX NCBI_TaxID=4097;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LEAF;
```

RX PubMed=10945337;
RA Hara K., Yagi M., Kozumi N., Kusano T., Sano H.;
RT "Screening of wound-responsive genes identifies an immediate-early
RT expressed gene encoding a highly charged protein in mechanically
RT wounded tobacco plants";
RL Plant Cell Physiol. 41:684-691(2000).
DR EMBL; AB009883; BAA95789.1; -
SQ SEQUENCE 513 AA; 61019 MW; F9598A37A840B970 CRC64;

alignment_scores:
Quality: 101.50 Length: 153
Ratio: 1.068 Gaps: 7
Percent Similarity: 62.092 Percent Identity: 21.569

alignment_block:

US-09-823-101-4 x Q9LW95 ..

Align seg 1/1 to: Q9LW95 from: 1 to: 513

```
54 CAAGAGCAAAAAAACCCGGGGAACCCGCGGGAACCCAG 103
|||||  |||||||  |||||||  |||||||  |||||||
178 GluLysAspLysLysGluLysAsnLysAsnLysGluLysGlyLysSerLy 194
104 GGAACATAGCGAGTGTCCCGGCTGGAACATGTGTACCCGCTACAAAA 153
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||
194 sGlyLysSerGlu.....GluGluT 201
154 CTCCAAAAGAG.....GAACATTGTCAACGCGCTCCACTCCCGG 194
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||
201 hrGluGluGluLysAspAspGluLysGlyLysAsnLysGluSerAspGlu 217
195 CACACAGCTAGAAACATCTACTAGAACTCACAACATATATACATA 244
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||
218 GluAspGluArgGluThrGlu....GluGluGluAsnAspGluLysGlu 232
245 GTACAGAGAAACGCCACATAGAGCTACTCGCATTAAGTACACTGCCGA 293
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
232 yValLysLysAspLysGluLysAsnLysGluLysGlyLysGluLys 249
294 .....AAGACACTATCAGAGACTGACGAGAGCAAGAA 326
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
249 spAsnGluLysLysAspLysSerLys.GluGluThrGluGluGluLysAs 265
327 CGACGACAGAGAGAACCAACAGACCCAGCGATGAG...CGCGATGAG 373
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||
265 pAspGluLysGluLysLysAspLysGluLysLysGlyLysLysAsnL 282
374 CAGACCGCAGATACCCGAAAGAACGGGAAAGACAGATGAAGTAAATC 423
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
282 yLysGluLysAspLysGluThr...LysAspLysSerLysGluVal 297
424 TCAGTTGACGAGAACCAAAAGCGAGACAGATGTTGATCAAGAAACCA 473
|||  |||  |||  |||  |||  |||  |||  |||  |||  |||
298 SerAspGluGluGluLysAspAspGluGluGlyGluLysAspLy 314
474 GCGCGAG 480
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
314 sLysLys 316

seq_name: sp_plant:Q49677

seq_documentation_block:
ID Q49677 PRELIMINARY; PRT; 1260 AA.
AC Q49677;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE PUTATIVE PROTEIN (PUTATIVE PROTEIN).
GN T18B16.160 OR ATAG19190.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
```

OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Bevan M., Benes V., Rechmann S., Borkova D., Ansoyge W., Bancroft I.,
 RA Mewes H.W., Mayer K., Schueller C.,
 RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Benes V., Rechmann S., Borkova D., Ansoyge W., Mewes H.W., Lemcke K.,
 RA Mayer K.F.X.,
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: BELONGS TO ZN-FINGER CCH TYPE FAMILY.
 DR EMBL: AL021687; CA16708.1; -;
 DR EMBL: AL161550; CAB7892.1; -;
 DR InterPro: IPR002885; PPR.
 DR InterPro: IPR001878; znf_CCHC.
 DR Pfam: PF01535; PPR; 8.
 DR Pfam: PF00098; zf-CCHC; 1.
 DR SMART: SM00343; znf_C2HC; 1.
 DR Zinc Finger.
 KM
 SQ SEQUENCE 1260 AA; 143517 MW; 626AC4A2F03F02B CRC64;

alignment_scores:
 Quality: 100.50 Length: 176
 Ratio: 1.047 Gaps: 7
 Percent Similarity: 54.545 Percent Identity: 24.432

alignment_block:
 US-09-823-101-4 x 049677 ..

Align seg 1/1 to: 049677 from: 1 to: 1260

```

31 TCAGGAAAAAACGTAATTCATCAAGACGAAAAAAGAAAAA 80
   ||| :|||:|||||:|||||:|||||:|||||:|||||:
360 SerArgLulysArgLysSerLysArgLysLysLysLysSe 376
   |||:|||||:|||||:|||||:|||||:|||||:
81 AAAAAACCCGGGGGAAACGAGGACATAGCGAGTCCCGCCTGA 130
   :|||:|||||:|||||:|||||:|||||:
376 rLysLysGln.....TyrAspSerAspSerLeuSerPheGlnGly.... 389
   :|||:|||||:|||||:|||||:|||||:
131 AACTGTGATACCGCGGTACAAACTCCAAAGAGAAACATTTCACACGC 180
   :|||:|||||:|||||:|||||:|||||:
390 ..SerGlySerAspSerTyrArgLeuSerArgArgHISThrLysHis 405
   :|||:|||||:|||||:|||||:|||||:
181 GCTCCACTACCGCGGCACACAGTA..... 204
   ||| :|||:|||||:|||||:|||||:
406 ValAsp...ProSerAlaSerLeuLysSerGlnValTyrHisGlnLys 421
   ||| :|||:|||||:|||||:|||||:
205 .....GAGAAACACATCTACTAGAAACTCACAACATATATACATA 244
   |||:|||||:|||||:|||||:|||||:
421 nSerHisArgGlnLysHisTyrTyrAspGlnLys.....HisG 434
   |||:|||||:|||||:|||||:|||||:
245 GTACAGAGAAACGCCACATAGACCTACTCGCAATAAGTCAACTCCGCCAA 294
   :|||:|||||:|||||:|||||:|||||:
434 LulysArgLysGlnLysValAspArgProSerAlaSerAspAspSer 450
   :|||:|||||:|||||:|||||:|||||:
295 AGACAGCGTTCAGACAGTACGACGAGAGAAAGACGACGAGAGACAA 344
   :|||:|||||:|||||:|||||:|||||:
451 AspTyrTyrArgSerAsnSerSerArgLysLysArgSerGlnAspSry 467
   :|||:|||||:|||||:|||||:|||||:
345 ACCAAGCAGCCAGCGTAGCGC..... 366
   :|||:|||||:|||||:|||||:|||||:
467 rLysSerHisArgGlnLysArgLysGlnValHisSerAsnAspProValS 484
   :|||:|||||:|||||:|||||:|||||:
367 ..GATAGGACGAGCGCAGATACCCAAAGAAAGACGGGAAAA.....AGA 408
   :|||:|||||:|||||:|||||:|||||:
484 erGlnLysSerGlnLysGlnHisTyrSerGlnSerGlnLysLysIleGlnArg 500

```

```

seq_name: sp_invertebrate:Q9W3D1
seq_documentation_block:
ID Q9W3D1 PRELIMINARY; PRT; 1183 AA.
AC Q9W3D1;
DT 01-MAY-2000 (TRENBLREL. 13, Created)
DT 01-MAY-2000 (TRENBLREL. 13, Last sequence update)
DT 01-JUN-2001 (TRENBLREL. 17, Last annotation update)
DE CG12109 PROTEIN (CHROMATIN ASSEMBLY FACTOR-1 P180 SUBUNIT).
GN CAF1-180 OR CG12109.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN-BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abtil J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.A., Bouck J., Brokstein P., Brotlier P.,
RA Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos L., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Dutilleul K.J., Evangelista C.C., Ferraz C., Fertiera S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,
RA Jaitai M., Katush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kuip D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Mounklov G., Mlshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skipski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Stokrop R., Sun E.,
RA Svirska R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
RA "The genome sequence of Drosophila melanogaster.",
RT Science 287:2185-2195(2000).
[2]
RP SEQUENCE FROM N.A.
RA Tyler J., Kadonaga J., Kobayashi R.,
RT "Drosophila CAF-1 subunits.",
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AE003444; AAF6399.1; -;
DR EMBL: AF367177; AAK31263.1; -;

```


DR FlyBase: FBgn0030054; Cafl-180.
SQ SEQUENCE 1183 AA: 133494 MW: 994DB576B401B9B CRC64:

alignment_scores:

Quality: 100.00 Length: 140
Ratio: 1.282 Gaps: 6
Percent Similarity: 55.714 Percent Identity: 26.429

alignment_block:
US-09-823-101-4 x Q9W3D1 ..

Align seg 1/1 to: Q9W3D1 from: 1 to: 1183

```
78 AAAAAAAAAACCGGGGGAACCAACGAGCATAGCGAGTCCCGGCT 127
   :::::::::::::: :::: :::: ::::
466 GlnylsGluThrAlaArgLeuAsnArgLysSerLeuProGluValThrGly.. 481
128 GGAACATGTGTACCGCGCTACAAACTCCAAAGAGAAACATTTGCAA 177
   :::: :::: :::: ::::
482 .....AlaProLysGlyLeuThrProLysGlnGlnArgLeuMetG 495
178 CGCGCTCCACACCGCGCACACAGCTAGAGAAACATCTTCTGAAACT 227
   :::: :::: :::: ::::
495 LcGlnAlaGlyLysAlaArgGluGluLysGluLysLeuAlaGluGlu 511
228 CACACATATATACATAGTACAGAAACCGCACATAGAGCTACTCGCAT 277
512 ArgAlaGlyLeuLys.....GlnGlnAspLysGluHisArgGluGln.. 525
278 AAAGTACAACTGCGCAAGACACGATATCAGAGCTGACGAGAGAAAG 327
   :::: :::: :::: ::::
526 .....LysLysGlnGluArgAspLysGluGlnGln 537
328 GACACAGAGAGACAAACCAACGACCGAGCT.....GAGCGCGCA 368
   :::: :::: :::: ::::
537 rGlyLsLeuGluArgAspLysGln..GlnGlnAlaGlyMetGluLysG 553
369 TGAGGACAGCGCGAGTACCGCAAGAAAGGGAACAGACAGATGAAG 418
553 uGluLysGluArg.....LysArgGlnAlaGluVal 563
419 TAATCTGAGTTGACGAAGACCAAGCAAGAGAGATGTGATCAAGA 468
   :::: :::: :::: ::::
563 AlaAspSerLysAsnGluGluLysArgLysArgAsnGluAlaLysGlu 579
469 AACCAAGCGCGACAGACAG 486
   :::: :::: ::::
580 ValGlnAlaGlyLysLysAsp 585
```

seq_name: sp_bacteria:087306

seq_documentation_block:

ID 087306 PRELIMINARY; PRT; 406 AA.
AC 087306;
DT 01-NOV-1998 (Tremblrel. 08, Created)
DT 01-NOV-1998 (Tremblrel. 08, last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, last annotation update)
DE ELPB2 (ORF29/ELPB2).
GN ORF29/ELPB2.
OS Borrelia burgdorferi (Lyme disease spirochete).
OC Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
OX NCBI_TaxID=139;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=297;
RA Akins D.R., Bourrell K., Calmano M.J., Hagman K., Norgard M.V.,
RA Radolf J.D.:
RT "An new animal model for studying Lyme disease spirochetes in the
RT host-adapted state".
RL Submitted (Sep-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.

RC PLASMID=CP18-2;
RX MEDLINE=20143776; PubMed=10678977;
RA Calmano M.J., Yang X., Popova T.G., Clawson M.L., Akins D.R.,
RA Norgard M.V., Radolf J.D.:
RT "Molecular and evolutionary characterization of the cp32/18 family of
RT supercoiled plasmids in Borrelia burgdorferi 297.";
RL Infect. Immun. 68:1574-1586(2000).
DR EMBL; AF023853; AAC34964.1; -;
DR EMBL; AF169008; AAF29801.1; -;
DR InterPro: IPR003483; OSPEF.
DR Pfam: PF02471; OSPEF; 1.
KW Plasmid.
SQ SEQUENCE 406 AA: 46543 MW: A3D92D3F42D3EB88 CRC64:

alignment_scores:

Quality: 99.00 Length: 140
Ratio: 1.207 Gaps: 5
Percent Similarity: 58.571 Percent Identity: 23.571

alignment_block:
US-09-823-101-4 x 087306 ..

Align seg 1/1 to: 087306 from: 1 to: 406

```
60 AAAAAAAAAAAAAAAAAAAAAAAAAACCGGGGGAACCAACGAGGACA 109
   :::::::::::::: :::: :::: ::::
142 GluGluLysGluGluLysGlnGluAsnAlaGluGluAsnThrLysGly.. 157
110 TAGCGAGTGTCCCGCGCTGAAACTGTGTACCGCGCTACAAACTCCA 159
   :::: :::: :::: ::::
158 .....LysGluValIleGluGlnGlnLysGln 168
160 AAGAGGAACATGTGCAACGCGCTCCACTACCGCGCAACAGCTAGAGAA 209
   :::::::::::::: :::: :::: ::::
168 LngLgLuThrAlaLysLysAla.....LysAlaGlnLysGluL 181
210 ACACATCTACTAGAAACTCACACATATATATAGTACAGAGAAACGCC 259
   :::: :::: :::: ::::
181 S.....ArgGluArg 185
260 ACATAGACTACTCGCATTAAGTACAACTGCGCAAGACAGCTATCAGAG 309
   :::: :::: :::: ::::
185 LngLgLuIleGlnGlnLysGlnGlnGlnGlnArgAlaLysGlu 201
310 ACTGACGAGAGAGAACGACGACGAGAGAGAGAACCAACGACGACG 359
   :::::::::::::: :::: :::: ::::
202 GluGluGlnGlnGlnArgArgAlaLysGluGlnGlnGlnArgAr 218
360 TGAGCGCATGAGCGCAGAG...CGCAGATACCGCAAGAACGCGGAGAA 406
   :::: :::: :::: ::::
218 GluLysGluGlnGlnGlnGlnArgArgAlaLysGluGlnGlnGlu 235
407 GACAGATGAAGTAACTCACTGACGAGAGCAACCAACGAGAGACAGAT 456
   :::: :::: :::: ::::
235 rGlnAlaArgArgAlaLysGluGlnGlnGlnGlnArgArgAlaLysGlu 251
457 GTTGATCAAGAAACGAG 474
   :::: :::: ::::
252 GluGluGlnLysArgGln 257
```

seq_name: sp_plant:Q9LH98

seq_documentation_block:

ID Q9LH98 PRELIMINARY; PRT; 2081 AA.
AC Q9LH98;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, last sequence update)
DT 01-OCT-2000 (Tremblrel. 15, last annotation update)
DE GENOMIC DNA, CHROMOSOME 3, BAC CLONE: T1998.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBITaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-COLUMBIA;
RA Kaneko T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.;
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-COLUMBIA;
RX PubMed=10907853;
RA Nakamura Y.;
RT "Structural analysis of Arabidopsis thaliana chromosome 3. II.
RT Sequence features of the regions of 4,251,695 bp covered by ninety P1,
RT TAC and BAC clones."
RL DNA Res. 7:217-221(2000).
DR EMBL: AP002057; BAB03174.1; -;
SQ SEQUENCE 2081 AA; 232852 MW; D3603E1F85EFF29 CRC64;

alignment_scores:
Quality: 99.00 Length: 160
Ratio: 1.151 Gaps: 5
Percent Similarity: 53.750 Percent Identity: 21.875

alignment_block:

US-09-823-101-4 x Q9LH98 ..

Align seg 1/1 to: Q9LH98 from: 1 to: 2081

```
37 AAAAAACGTAATCTATCAAGAGAAAAA.....ATAGCAGTGTCCCG 124
|||||.....|
1213 LysLysGlnThrSerValGluGluAsnLysLysGlnLysGlnThrLysLys 1229
87 ACCCGGGGGGAACGAGGGGAC.....ATAGCAGTGTCCCG 124
|||||.....|
1229 sGluLysAsnLysProLysAspLysLysAsnThrThrLysGlnSerG 1246
125 GCTGAAGAACTGTGTACCCGCTACAAAACTCCAAAAGAGAAACATTGT 174
|||.....|
1246 LysLysLysLysGlnSerMetGlnSerLysGlnLysGlnLysGln 1262
175 CAACGCGCTCCACTACCGCGCAGACAGCTAGAGAAACACATCTAGTAA 224
|||||.....|
1263 GlnLysSer.....Gln 1266
225 ACACAGCAACATATATACATAGTACAGAGAAACCCACATAGACTACTCG 274
:|||||:|
1266 nAlaThrThrGlnAlaAspSerAspLysLysAsnGlnLysLeuMetG 1283
275 CATTAAGTCAACTCGCAAGACAC.....GTATCAGAGACTGAC 315
:|||||:|
1283 LalaAspSerGlnAlaAspSerHisSerAspSerGlnAlaAspSerAsp 1299
316 GAGAGAGAAAAAGAC.....GACAGAGGAGACAAACCAAGAGCCA 356
|||.....|
1300 GluSerLysAsnGlnLysLeuMetGlnAlaAspSerGlnAlaThrThrG 1316
357 GCGTGAGCGGATGAGGACAGCGCAGATACCCGAAAGAAACGGGAGAAA 406
|||||.....|
1316 nArgAsnAsnGlnLysAspArgLys..... 1324
407 GACAGATGAAGTAATCTCAGTTCAGACAGAGCAAGCAAAAGCAGAGAGAT 456
|||.....|
1325 .....LysGlnThrSerValAlaGluAsnLysLysGlnLysGlnThr 1338
457 GTTGATCAAGAAACGAGCGGAGAGACAG 486
:|||||:|
1339 LysGlnGluLysAsnLysProLysAspAsp 1348
```

seq_name: sp_invertebrate:Q23319

seq_documentation_block:

ID Q23319 PRELIMINARY; PRT; 389 AA.
AC Q23319;
DE 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DI 01-JAN-1999 (TREMBLrel. 09, Last annotation update)
DE ZC43.4 PROTEIN.
GN ZC43.4.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBITaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Baynes C.;
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Alnscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favell A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,
RA Smaldon N., Smith A., Sonhammer E., Staden R., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sproat J., Woldman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans."
RL Nature 368:32-38(1994).
DR EMBL: Z75553; CA99946.1; -;
SQ SEQUENCE 389 AA; 44984 MW; 5812D07D30329297 CRC64;

alignment_scores:
Quality: 98.50 Length: 154
Ratio: 1.187 Gaps: 5
Percent Similarity: 53.896 Percent Identity: 24.026

alignment_block:

US-09-823-101-4 x Q23319 ..

Align seg 1/1 to: Q23319 from: 1 to: 389

```
58 AGSAAAAA.....AACCGGGGGAACCAAGCGGA 107
|||||.....|
162 ArgAlaLysSerArgSerGlnLysAsnValProAlaTyrSerPheAsp 178
108 CATAGCGAGTGTCCCGGC.....TGGAAC 133
| |||:|||||
178 pProAlaLeuLeuIleProGlyGluGlyProIleProGluArgAlaThrLysV 195
134 TGTGTACCCGGCTACAAACTCCAAAGAGAAACATTGTCAAGCGCT 183
:|||||:|
195 alTyrTyrAsnAlaValLysSerValGlnAlaProThrProGluGlnAla 211
184 CCACCTACCGCGCACACAGCTAGAGAAACACATCTACTAAGACTGCAC 233
|||||.....|
212 ProValAlaGlnAspValAlaValAsnGluLeu..... 222
234 ATATATACATAGTACAGAGAAACGCCACATAGACTACTCGCATTAAGTA 283
|||.....|
223 .....SerGlnValLysLysTyrGlyGlnLeu..... 232
284 CAATCGCGCAAGACACGATATCAGAGACTGACGAGAGAAACGAGAGAC 333
:|||||:|
233 .....LysAlaGlnMetGluGluAlaGluLysArgArgGluGluAsp 246
334 AGAGGAGACAAACCAAGCAGAGCGCTGAGCGGAGAGAGAGAGCGGACG 383
|||.....|
247 .....AspValAlaAsnLysLeuAsnGluValValGluLysLeuIleSe 261
```

```

384 ATACCCGAAGAAAGGGGAAAGACAGATGAAGTAATCTCACTTGACG 433
      |||:|||||
261 rThAlAlAspSerGlyLysSerGln.....LysG 272
434 AAGGCCAAAGACGAGAGCATGTTGATCAAGAAACCGACGAGACA 483
      |||:|||||
272 LuAspGluGluGluAsnGluAspAspAspAspAspGluGluGlu 288
484 GACTGTGCAGAA 495
      ::|||
289 GluAsnValGlu 292

seq_name: sp_virus:055708

seq_documentation_block:
ID 055708; PRELIMINARY; PRT; 456 AA.
AC 055708;
DT 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-JUN-1998 (TREMBLrel. 06, last sequence update)
DE 01-JUN-1998 (TREMBLrel. 06, last annotation update)
DE HYPOHETICAL 52.8 KDA PROTEIN.
OS Chilo Iridescens virus (CIV) (insect Iridescens virus type 6).
OC Viruses; dsDNA viruses, no RNA stage; Iridoviridae; Iridovirus.
OX NCBI_TaxID=10488;
RN [1]
RP SEQUENCE FROM N.A.
RA Bahr U., Tidona C.A., Darai G.;
RL Virus Genes 0:0-0(1997).
DR EMBL: AF003534; AAB94419.1; -.
KW Hypoetical protein.
SQ SEQUENCE 456 AA; 52815 MW; 290CD897A855B798 CRC64;

```

```

alignment_scores:
Quality: 98.50 Length: 169
Ratio: 1.107 Gaps: 9
Percent Similarity: 52.663 Percent Identity: 26.036

```

```

alignment_block:
US-09-823-101-4 x 055708 ..

```

```

Align seg 1/1 to: 055708 from: 1 to: 456

```

```

51 TATCAAGAGCAAAAAAAAAAAAAAAAAAACCGGGGGAAC 100
    ||| ||| :|||:||||| |||
30 TyrlGluGluLeuGluAsnLysValLysLeu..... 42
101 CAGGGACATAGCGAGTGTCCCGGCTGGAACCTGTG....GTACCG 144
    :||| ||| ||| |||
43 .....AsnGlyLeuGluPheValProThrProP 55
145 GCTCAAAACCTCCAAAGAGGAACATTGTCA.....ACG 179
    :||| ||| ||| |||
55 rGluSerThrGlyArgGluThrLeuValSerLysAspProHisValVal 71
180 CGCTCCACTACCGGACACAGCTAGAGAAACATCTACTAGAACTCA 229
    :|||:||||| :|||:|||||
72 GlnSerSerLysSerAsnThrProLeuGhrGluLeuLysAspThrPr 88
230 CAACATATATACATAGTACAGAGAAACCCACATAGACTACTCGCATTA 279
    :|||:||||| |||:|||||
88 OArg.....TyrGluGluThrProLeuLysArgThrLeuThrV 101
280 AGTCAACTGC.....GCA 293
101 alThrnValLysThrValLysSerSerSerLysGlyMetAsnGly 117
294 AAGACACGTATCAGAGACTGACGAGAGAAGACGACGAGAGAGACA 343
    :|||:||||| :|||:|||||
118 ArgAsnArgLeuTyrAsp.AspsAspLeuPheAspAspArgTyrLys 134
344 AACCAAGACGCGACGCT.....GAGCGGATGAGACGAGCGC 381
    :|||:||||| |||

```

```

134 ePrThrThrArgLysPheGlnGlyLysArgAspGluAspLleArg 150
382 AGATACCCGAAGAAACG.....CGGAAAGACAGATGAAGTAATC.. 423
    |||:||||| :|||:|||||
151 LeuLeuProLysSerSerAsnLleGlySerSerLysTyrLysProValLe 167
424 ....TCAGTTCAGACAGACCAAGACGAGAGACGAGTGTGATCAAGAA 469
    |||:||||| :|||:|||||
167 uThrArgValGluGluAsnGluAsnLysLysLleHisLleAspGlnArgL 184
470 ACCAG 474
184 ysglu 185

```

```

seq_name: sp_invertebrate:017840

```

```

seq_documentation_block:
ID 017840; PRELIMINARY; PRT; 490 AA.
AC 017840;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, last sequence update)
DE 01-JUN-2001 (TREMBLrel. 17, last annotation update)
DE P25H5.7 PROTEIN (EC 3.1.3.48).
GN P25H5.7.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Steward C.;
RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
RN [2]

```

```

RP MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rikken L., Roopra A., Saunders D., Showkeen R.,
RA Smaildon N., Smith A., Sonhammer E., Staden K., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sproat J., Woldman P.;
RT *2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.
RL Nature 368:32-38(1994).
DR EMBL: Z81068; CAB02988.1; -.
DR HSSP: P18052; IYFO.
DR InterPro: IPR000387; TYR_phosphatase.
DR InterPro: IPR000242; Tyr_prot_phphatase.
DR Pfam: PF00102; Y_phosphatase.
DR PRINTS: PR00700; PRTYPHPTASE.
DR SMART: SM00194; PTPC; 1.
DR PROSITE: PS00383; TYR_PHOSPHATASE_1; 1.
DR PROSITE: PS50056; TYR_PHOSPHATASE_2; 1.
DR PROSITE: PS50055; TYR_PHOSPHATASE_PTP; 1.
KW Hydrolase.
SQ SEQUENCE 490 AA; 56426 MW; 94FA5DA4F3531644 CRC64;

```

```

alignment_scores:
Quality: 98.50 Length: 132
Ratio: 1.201 Gaps: 6
Percent Similarity: 62.121 Percent Identity: 28.030

```

```

alignment_block:
US-09-823-101-4 x 017840 ..

```

```

Align seg 1/1 to: 017840 from: 1 to: 490

```

```

69 AAAAAAAAAAAAAAAAAAACCGGGGGAACACAGGACATAGCGAGTG 118

```

THIS PAGE BLANK (USPTO)

156 CTGCGCTCTCTGCGCTACCTTTCCACGACGACCTCTCTTTTCT 107
 164ValThrThrIlePheIle 169
 106 CTTTTCCTATTGTCTACGCCCTTCCCTCCTACTTAC..... 68
 170 ILeYrAlaPheLeuAsnTyrProPheGlyIuAsnAsnLysAsnAspVa 186
 67CTTAGAGTCTCGCTCTGTATACCATATGTGAGCTTCA 28
 166 lCysTyrThrLeuArgPheSerAsnGlyMetGluLeuValArgIlePheS 203
 27 CTCGAGATC 20
 203 eTThrIleu 205

seq_name: p1r2:A36886

seq_documentation_block:
 surface protein P4g negative regulator par - Streptococcus sobrinus (strain MT3791)
 C:Species: Streptococcus sobrinus
 C:Date: 04-Nov-1994 #sequence_revision 04-Nov-1994 #text_change 20-Jun-2000
 C:Accession: A36886
 R:Takahashi, I.; Okahashi, N.; Hamada, S.
 J. Bacteriol. 175, 4345-4353, 1993
 A:Title: Molecular characterization of a negative regulator of Streptococcus sobrinus su
 A:Reference number: A36886; MUID:933232312
 A:Accession: A36886
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-203 <TAK>
 A:Cross-references: GB:D13323; NID:g217037; PIDN:BAA02581.1; PID:g425488
 A:Note: It is uncertain whether Met-1 or Met-7 is the initiator
 C:Genetics:
 A:Gene: par
 A:Start codon: GTG
 C:Superfamily: Streptococcus sobrinus surface protein P4g negative regulator par
 C:Keywords: DNA binding

alignment_scores:
 Quality: 80.50 Length: 181
 Ratio: 1.134 Gaps: 10
 Percent Similarity: 39.227 Percent Identity: 25.414

alignment_block:
 US-09-823-101-1/rev x A36886 ..

Align seg 1/1 to: A36886 from: 1 to: 203

532 TGTTCCTCCCTCTTTTCTTTT..... 509
 33 CysIleProPhePhePheSerTrpLeuSerArgLeuThrIleG1 49
 508TTTTTTTTTTGG.....TTTTTTTAACTT 481
 49 yIleSerAsnIlePhePheAsnTyrTrpIleuProPhePheIleA 66
 480 GAGGTAGTTTCTTTAGTACTGTGTCCATATGCGCATATTTCG 431
 66 heAlaLeuAlaLeuLeuLeuValIle.....HisGluGlyIle..... 78
 430 AACTTGGCAGCAGGAATTAAACCAATTTTTGTAGTCAGAAATTC 381
 78 78
 380 AAAAGTATTCTGTCTCTTATATAAATTGGAATATATAAAGAACAA 331
 78 78
 330 GAATCAGCATATTCACACATCAGACAAATCAGTGCATATTTCG 281
 79HisGlyPheTyrPheIys 84

280 ATATT...CCTTCCTGCTCTTTTCTCTGCAATATAGAAAACGAGAGC 234
 85 LeuPheLysProGluAsnProLeuLys...TyrGlyThrAspTrpArgLe 100
 233 TGGATTTTCAGTGGCAGTGCAGCATGAGTGCACCTACACAGAAATGTC 184
 100 uGlyLeuPheAsnAlaThrSerProGlySerArgTyrProArg..... 114
 183 GTTTCATATTAAGGTCATGTCGTGCTGCTGCTGCTGCTGCTGCTGCT 134
 115SerGluMetLeuIleIleTyrIleu 122
 133 TCCCA...CTAAGACCTCTCTTTTCTCTCTCTCTCTCTCTCTCTCT 91
 123 AlaProPheValIleuThrSerLeuLeuThrLeuLeuAlaLeuG 139
 90CTAGCCCTTCCCTCCTACTTAACCTTAGAGATCTG 55
 139 lYrThrLeuSerProLeuAlaTyrIleuPheLeuAlaValIle 152

seq_name: p1r2:D83867

seq_documentation_block:
 hypothetical protein BHI740 [imported] - Bacillus halodurans (strain C-125)
 C:Species: Bacillus halodurans
 C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 31-Dec-2000
 C:Accession: D83867
 R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fuji, F.; H
 Nucleic Acids Res. 28, 4311-4331, 2000
 A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans a
 A:Reference number: A83650; MUID:20263314
 A:Accession: D83867
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-457 <STO>
 A:Cross-references: GB:AP001513; GB:BA000004; NID:g10174345; PIDN:BA05459.1; GSPDB:G
 A:Experimental source: strain C-125
 C:Genetics:
 A:Gene: BHI740

alignment_scores:
 Quality: 80.50 Length: 195
 Ratio: 0.885 Gaps: 9
 Percent Similarity: 46.667 Percent Identity: 24.615

alignment_block:
 US-09-823-101-1/rev x D83867 ..

Align seg 1/1 to: D83867 from: 1 to: 457

505 TTTTGTGTTGTTTAACTGAGTAGTGTCTTCTTTAGTAT 456
 67 PheCysPheTrpIleGluGlyAlaSerGlySer.....ThrIle 79
 455 CTGTGTCATTAATGACATATTTGAACTGGCAGGGAATATAAAGAAC 406
 79 uValPheSerPheGlyGlnAsnLeuAsnIleIleSer..... 91
 405 CAATTTTGTAGTCAGAAATTTCTTAAAGTATTTCTGCTCTTATATA 356
 92 ..MetPheLeuPheIleProPheIleSerThrPheLeuSerSerValGly 107
 355 AATTGGAAATATATAAAGAA.....CAAGAATACAGCATTAATTCAC 312
 108 TyrIleAspSerMetLysHisTrpValGlnLysGlnGluLysArgAr 124
 311 ACTCAGACAAATCAGTGCATATTTGGCATATTTC..... 273
 124 gLeuArgProAspHisThr..AlaPheTrpLeuThrSerIleValGlyLe 140
 272 TTCCGTGCTCTT..... 261

```

140 uleuLeuAsnPhleGlySerLeuAlaIleValLysArgIleIleGlyLys 157
    ::::::::::::::
260 ..TTCTGTGATATAGAAAAGCTGGAGAGCTGGATTTTTCAGTGGCAGTGA 213
    |||||:::|||||:
157 erPheserSerPhearGlu..... 163
212 GCATGAGTGGACCTACACAGAGATTTCTGTTTCCATATTAAAGTCATG 163
    ::::::::::::::
164 .....GlnArgLeuMetLeuValIleLeuArgGlyPh 174
162 TGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 123
    ||| |||||::: |||||::: |||||
174 egLySerCysLeuLeuTrpSer...ProPhemetValAsnIleGlyLeu 190
122 .....GACCTCTCTTTTCTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 87
    |||||:::|||||:
190 leuThrIlePhesPleuSerTrpTyrGlnIleGlyGlyPhGlyLeu 206
86 GCCCTTCCCTCCTCACTTAACCTTAGAGTCTGCTGCTGCTGCTGCTGCT 37
    :: ||| ||||| ||||| |||||
207 leuMetAlaLeuIleTyrIlePhelLeuPhetTrpLeuPhetTyrProAla 223
36 GAGCTCTCAGCTGAGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3
    ::::: ||| ::::: |||
223 egLInPhesPHisAspProValIleGluHisHis 234

seq_name: p1r2:A45989

```

```

seq documentation block:
peroxisome assembly factor-1 - Chinese hamster
C:Species: Cricetus griseus (Chinese hamster)
C:Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 20-Jun-2000
C:Accession: A45989; A53782; S31621
R:Thieringer, R.; Raetz, C.R.H.
J. Biol. Chem. 268, 12631-12636, 1993
A:Title: Peroxisome-deficient Chinese hamster ovary cells with point mutations in peroxi
A:Reference number: A45989; MUID:93286102
A:Accession: A45989
A:Molecule type: mRNA
A:Residues: 1-304 <THI>
A:Cross-references: GB:217220; NID:949474; PIDN:CAAT8929.1; PID:949475
R:Tsukamoto, T.; Shimozawa, N.; Fujiki, Y.
Mol. Cell. Biol. 14, 5458-5465, 1994
A:Title: Peroxisome assembly factor 1: nonsense mutation in a peroxisome-deficient Chine
A:Reference number: A53782; MUID:94309666
A:Accession: A53782
A:Molecule type: mRNA
A:Residues: 1-304 <TSU>
A:Cross-references: GB:D30618; NID:9531027; PIDN:BAA06308.1; PID:91109617
C:Superfamily: peroxisome assembly factor-1
C:Keywords: transmembrane protein

```

```

alignment_scores:
  Quality: 80.00      Length: 171
  Ratio: 0.930      Gaps: 8
  Percent Similarity: 50.292      Percent Identity: 22.807

```

alignment block:

US-09-823-101-1/rev x A45989 ..

Align seg 1/1 to: A45989 from: 1 to: 304

```

478 GGTAGTTTCTCTTTAGTATCTGTTCCATTAATGATATTTCTGAA 429
    ||:::|||||: |||||
148 GtGtLleuLeuLeuAsnPhleLeuIlePhelGlnLysGlyLysPhelAla 164
428 CTGGCAGAG.....GGAATTAATAAACCAATTTTCTAGTCAGA 388
    |||||::: |||||:::
164 leuThrGluArgLeuLeuGlyLysHisSer...ValPhesCysLysProG 180
387 AATTTTAAGATATTTCTGTGCTCTTATTAATAATTTGCAATATATAAAA 338

```

```

180 InAsnIleArgGluValGlyPhesPTrpMetAsnArgGluLeu..... 194
337 GAACAAGAAATCAGCAGCAATTAATTAACCACTGACAGCAAAATCAGTGA 288
195 .....LeuTrpHisGlyPheAlaGluPh 202
287 TTTTGCAATATTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 256
    :||| ::::: |||||:::
202 eLeuIlePhelLeuLeuProleuLeuAsnIleGlnLysPhelAlaLys 218
255 .....TCATATATAGAAAAGCTGGAGAGCTGGATTTTTCAGTGGC 218
    ||||| ||||| |||||
219 leuSerSerTrpCysIleProleuThrGlyAlaAlaSerSerAspSerAl 235
217 ACTAGAGATGAGTGGACCTACACAAAGATGTTGCTTTTCATATTAAAG 168
235 aleuAlaSerSerGly.....LysGluCys..... 243
167 TCATGTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 118
    ::||| :: |||||::: |||||
244 .....AlaLeuCysGlyGluTrpProThrMetProHisThrIleGly 257
117 CTCTTTTCTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 68
    ::||| ::||| ::|||
258 CysGluHisValPhesCysTyrTyrCysValLysSerSer.....PheLe 272
67 CTTAGAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 18
272 upPhesPMetTrpPhetTrpCysProLysCysGlyIleGluValHisSer 289
17 TGGTTCCAATG 7
289 alGlnProLeu 292

seq_name: p1r2:C84914

```

```

seq documentation block:
hypothetical protein At2g47360 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: C84914
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon,
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter
Nature 402, 761-766, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487
A:Accession: C84914
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-303 <STO>
A:Cross-references: GB:AE002093; NID:92275214; PIDN:AAB63836.1; GSPDB:GN00139
C:Genetics:
A:Gene: At2g47360
A:Map position: 2

```

```

alignment_scores:
  Quality: 79.50      Length: 91
  Ratio: 1.728      Gaps: 3
  Percent Similarity: 50.549      Percent Identity: 28.571

```

alignment block:

US-09-823-101-1/rev x C84914 ..

Align seg 1/1 to: C84914 from: 1 to: 303

```

348 AATATTAATAAGAAAGAAATCAGCAGCAATTAATTAATTAATTAATTAAT 299
    ||:::|||||: |||||
7 LysTrpArgLysGluAlaGluGlnLeuValValLysProPhesArgLeuVa 23
298 CACTGTGACTATTTTGGCAATATTTCTGCTGCTGCTGCTGCTGCTGCTAT 249

```

```

:||||:||||:||||: ||| ||| |||
23 1ThrhThrhLeuLeuSerLeuLeuProLeuSerPheLeuLeuLeu. 39
248 AGAAACTGAGAGCTGGATTTCAGTGCCACTGACGATGATGAGCACT 199
40 SerArgLeuSerSerAlaSerPhe..... 47
198 ACACAGAAATGTCGTTCATATTAAGCAGATGTCGTCTCCCTGC 149
48 .....LeuPheSerLeuThrlYsSer..... 54
148 TCCTGCGCCTACCTTCCCATATGACCTCTCTTTTTCCTTTTCC 99
55 .....GlnProGlnThrlGlnSerSerPheValPheSer 66
98 TATTTGTTCAGCCCTTCC 78
67 LeuPheLeuArgAlaAsnPro 73
seq_name: pir2:S46187
seq_documentation_block:
probable membrane protein YKL219w homolog YBR302c - yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein YBR2121; hypothetical protein YM4987.03; hypothe
C:Species: Saccharomyces cerevisiae
C>Date: 26-Aug-1994 #sequence_revision 09-Sep-1994 #text_change 24-Sep-1999
C:Accession: S46187; S46186; S55818; S58196
R:Feldmann, H.; Manhaupt, G.; Schwarzlose, C.; Vetter, I.
submitted to the Protein Sequence Database, August 1994
A:Reference number: S45927
A:Accession: S46187
A:Molecule type: DNA
A:Residues: 1-47 <FEU>
A:Cross-references: EMBL:Z36170; MIPS:YBR302c
A:Experimental source: strain S288C
A:Genetics: C02
R:Feuermann, M.; Potier, S.; Souciet, J.L.
submitted to the Protein Sequence Database, August 1994
A:Reference number: S46181
A:Accession: S46186
A:Molecule type: DNA
A:Residues: 30-379 <FEU>
A:Cross-references: EMBL:Z36170; MIPS:YBR302c
A:Experimental source: Strain S288C
A:Genetics: C02
R:Feuermann, M.; Charbonnel, L.; de Montigny, J.; Bloch, J.C.; Potier, S.; Souciet, J.L.
Yeast 11, 667-672, 1995
A:Title: Sequence of a 9.8 kb segment of yeast chromosome II including the three genes C
A:Reference number: S55813; MUID:56093905
A:Accession: S55818
A>Status: translation not shown
A:Molecule type: DNA
A:Residues: 30-379 <FEU>
A:Cross-references: EMBL:Z36171
A:Genetics: C02
R:Bowman, S.
submitted to the EMBL Data Library, July 1995
A:Reference number: S58194
A:Accession: S58196
A:Molecule type: DNA
A:Residues: 1-379 <BOW>
A:Cross-references: EMBL:Z50178; MID:9927528; PID:CAA90551.1; PID:9927530; MIPS:YML132w
A:Experimental source: strain AB972
A:Genetics: C13
C:Genetics: <C02>
A:Map position: 2R
A:Note: YBR302c
A:Molecule type: DNA
A:Residues: 1-379 <BOW>
A:Cross-references: MIPS:YML132w; SGD:S0004601
A:Map position: 13L
A:Superfamily: conserved hypothetical protein YKL219w
A:Keywords: transmembrane protein

```

```

Fj:46-62/Domain: transmembrane #status predicted <TM1>
F:65-96/Domain: transmembrane #status predicted <TM2>
F:233-253/Domain: transmembrane #status predicted <TM3>
F:254-273/Domain: transmembrane #status predicted <TM4>

alignment_block:
    alignment_scores:
        Quality: 79.50          Length: 128
        Ratio:   1.187         Gaps:   6
        Percent Similarity: 52.344      Percent Identity: 23.438
    align seg I/I to: S46I87 from: 1 to: 379
..
seq_name: pIR2:S69731
seq_documentation_block:
    hypothetical protein YDR452w - yeast (Saccharomyces cerevisiae)
C:Species: Saccharomycetes cerevisiae
C>Date: 22-Aug-1996 #sequence_revision 06-Sep-1996 #text_change 23-Mar-2001
C:Accession: S69731
R:Dietrich, F.S.
submitted to the EMBL Data Library, August 1995
A:Description: The sequence of S. cerevisiae lambda 3641 and cosmids 9461, 9831, and
A:Reference number: S69555
A:Accession: S69731
A:Molecule type: DNA
A:Residues: 1-674 <DIE>
A:Cross-references: EMBL:U03007; NID:g927685; PID:g927706; GSPDB:GN00004; MIPS:YDR452w
C:Genetics:
A:Gene: MIPS:YDR452w
A:Map position: 4R

alignment_scores:
    quality: 79.50          length: 110
    ratio:   1.242         gaps:   6
    percent similarity: 58.182     percent identity: 29.091
```

alignment_block:
US-09-823-101-1 x S69731 ..

Align seg 1/1 to: S69731 from: 1 to: 674

```

256 GAGAAAAAGCAGACGAGAA.....ATATGCCAAATAGTCACAG 296
    |||||||.....|
463 GlutylsAspLysLysLysLysProLethrArgLysGluLeu1 479
    |||||||.....|
297 TGATTTGTCGAGTGTGAATTATGCGTATCTTGTCTTTTATAT 346
    :|||:|:|:|
479 egluAaGTYSerLileValasnLilegLYSerValIleproThrPhea 496
    :|||:|:|:|
347 TTCCAAAATTT.....TATAAGACACAGAAATACCT..... 378
    |||:|:|:|
496 snProSerPheArgIleTrpGluTYrAsnLleThraSpLileValasnasp 512
    |||:|:|:|
379 .....TTAGAAATTCGACTACAAAAATTCG.....TTTTTAATT 416
    :|:|:|:|:|:|
513 SerAsnPheAlaValSerGluTYrLysPro.TTPAspGluPheGluS 529
    :|:|:|:|:|:|
417 CCCTGTGCCAAGTTCAGAAATATGCCATATGACACAGAT..... 457
    |||||:|:|:|:|:|:|
529 etLeuasnLysIleMetGluAspSerLeuLeuGluAspGluMetAspSer 545
    :|:|:|:|:|:|
458 ...AACTAAAGAGAAAGAACTACCTCAAGGTTAAAAACCAAAAAAAA 504
    |||:|:|:|
546 SerAsnIleGluValGlyIleAsnArgGluLysMetGlyGluLysLysAs 562
    :|:|:|:|
505 AAAAAAAAAAAAAAGGGGGGGAACA 532
    :|:|:|:|:|:|:|:|:|:|:|:|
562 nLysLysLysLysLysAsnAspLysThr 571

```

seq_name: plr2:T21884

seq_documentation_block:

hypothetical protein F36H1.2 - *Caenorhabditis elegans*
 C:Species: *Caenorhabditis elegans*
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
 C:Accession: T21884
 R:Kershaw, J.
 submitted to the EMBL Data Library, January 1996
 A:Reference number: Z19482
 A:Accession: T21884
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1398 <NTL>
 A:Cross-references: EMBL:Z68760; PIDN:CAA92996.1; GSPDB:GN00022; CESP:F36H1.2
 A:Experimental source: clone F36H1
 C:Genetics:
 A:Gene: CESP:F36H1.2
 A:Map position: 4
 A:Introns: 38/3; 60/3; 94/3; 152/3; 191/3; 257/2; 290/3; 323/3; 379/1; 439/1; 485/2; 552/2

alignment_scores:

Quality:	79.00	Length:	63
Ratio:	2.026	Gaps:	2
Percent Similarity:	61.905	Percent Identity:	33.333

alignment_block:

US-09-823-101-1/rev x T21884 ..

Align seg 1/1 to: T21884 from: 1 to: 1398

```

519 TTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 470
    |||||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
491 PhePheAlaIlePheLeuPheGlyMetPheSerLeuThrPheThrPhe 507
    :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
469 CTTCTTAGTATCTGTTCATATGCAATATTCGAACTTGTGCACAA 420
    :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
507 tLeuIleAlaIleSerAsnSerValThrAlaTYrLeuIleSerTrpServ 524

```

```

419 GCGAATTAAAAACCAATTTTGTACTCAGAAATCTTAAAGTATTTC 370
    |||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
524 alPheLeu.....tLeuIlePheIleIle.....Phe 532
    :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
369 TGTCCTCTTATAAATTTGGAATATATAAAGACACAA 331
    |||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
533 CysSerLeuIleValValTYrTYrGlyAspArgLys 545

```

seq_name: plr2:D81387

seq_documentation_block:

probable integral membrane protein Cj0430 [imported] - *Campylobacter jejuni* (strain N)
 C:Species: *Campylobacter jejuni*
 C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 31-Mar-2000
 C:Accession: D81387
 R:Parkhill, J.; Wren, B.W.; Mungall, K.; Kelley, J.M.; Churcher, C.; Basham, D.; Chli, C.W.; Quail, M.; Rajandream, K.A.; Rutherford, K.M.; Vanylic, A.; Whitehead, S.; Ba, Nature 403, 665-668, 2000
 A:Title: The genome sequence of the food-borne pathogen *Campylobacter jejuni* reveals
 A:Reference number: AB1250; MUID:20150912
 A:Accession: D81387
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-408 <PAR>
 A:Cross-references: GB:AL139075; GB:AL111168; NID:96967817; PIDN:CA874266.1; PID:9696
 A:Experimental source: serotype O2, strain NCTC 11168
 C:Genetics:
 A:Gene: Cj0430

alignment_scores:

Quality:	78.50	Length:	119
Ratio:	1.454 <td>Gaps:</td> <td>6 </td>	Gaps:	6
Percent Similarity:	45.378	Percent Identity:	26.050

alignment_block:

US-09-823-101-1/rev x D81387 ..

Align seg 1/1 to: D81387 from: 1 to: 408

```

541 TTTGGCCCTTGTTCCGCCCTTTTGTGTGTGTGTGTGTGTGTGT 497
    |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
212 PheAlaIalCysPheSerProLeuValPheValTYrPhePheTYrThr11 228
    :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
496 .....TGGTTTTTTTAAAC 484
    :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
228 eTYrArgLeuThrPheGlnLysTYrLysAsnLeuLeuThrPhe..... 242
    :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
483 CTTGAGTAGTATTCTTCTTTAGTTATCTGTTCATATGCAATATT 434
    |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
243 .....LeuMetSerValThrPheValAlPheCysLeuLeu 253
    :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
433 CTGAACTTGGCACAGGAATTAAAAACCA.....AT 402
    |||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
254 LeuSerLeuArgGlnLysLeuPheLeuAspAspPheLeuProPheCysVa 270
    :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
401 TTTTGTAGCAGAAATCTTAAAGTATTTCGCTCTTATAA..... 356
    :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
270 IileCysInProLeuLeuIleLysThrLeuMetGlnSerTYrArgValA 287
    :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
355 .....AATTGGAATATATAAAGAACAAAGATACACCAT 320
    :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
287 rglLeuLeuValPheArgLeu.ArgTYrLysIlePheIleGluCysSer11 303
    :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
319 AATTACCCACTCAGACAAATATCATCTGTGACTATTTGGCATATTCCTTC 270
    |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
303 eIle.....PheLeuIlePheCysTYrPheLeu1 313
    :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
269 CTGTC 265
    |||
313 leVal 314

```

seq_name: plr2:S64837

```

seq_documentation_block:
  hypothetical protein YLR015w - yeast (Saccharomyces cerevisiae)
  N:Alternate names: hypothetical protein L1583
  C:Species: Saccharomyces cerevisiae
  C>Date: 01-Aug-1995 #sequence_revision 24-May-1996 #text_change 29-Oct-1999
  C:Vandenbol, M.; Portetelle, D.; Hliger, F.
  submitted to the Protein Sequence Database, May 1996
  A:Reference number: S64742
  A:Accession: S64837
  A:Molecule type: DNA
  A:Residues: 1-505 <V>
  A:Cross-references: EMBL:Z7187; NID:q1360313; PIDN:CAA97537.1; PID:e245494; PID:g136031
  A:Experimental source: strain S288C
  R:Saville, S.P.; Atkinson, S.; Jamieson, L.; Pocklington, M.J.; Orr, E.
  submitted to the EMBL Data Library, August 1995
  A:Description: The 5.8kb fragment from chromosome XII of Saccharomyces cerevisiae does not
  A:Reference number: S59270
  A:Accession: S59274
  A:Molecule type: DNA
  A:Residues: 1-464, 'ISLL', 468 <SAV>
  A:Cross-references: EMBL:X90564; NID:g975221; PIDN:CAA62158.1; PID:g975226
  A:Experimental source: strain S288C
  R:Levin, D.E.; Stevenson, W.D.
  submitted to the EMBL Data Library, July 1994
  A:Description: The S. cerevisiae PKC2 does not exist in the yeast genome.
  A:Reference number: S48527
  A:Accession: S48527
  A:Molecule type: DNA
  A:Residues: 1-472 <LEV>
  A:Cross-references: EMBL:L34405
  C:Genetics:
  A:Map position: 12R

alignment_scores:
  Quality: 78.50      Length: 174
  Ratio: 0.882      Gaps: 7
  Percent Similarity: 51.149      Percent Identity: 24.138

alignment_block:
  US-09-823-101-1 x S64837 ..
  Align seg 1/1 to: S64837 from: 1 to: 505

80 GAAGGCTAGAACAAATAGAGAAAAAGAGAGAGAGCTCATTAATGTC 129
  ||| |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
  275 GIUASpAlaGluAnsnglyProSerArgLysLysLeuLysGlnGluThrTh 291
  130 GGGAAGGCTAGGCCAGAGAGA..... 151
  291 rAsnLysGluPhcGlnrGAlaLeuGluAnsprllegluTYrAsnAspy 308
  152 ..GGCAGACACACATGACCTTAATATGAAAAAGACAGACATTCCTGTGTA 199
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
  308 aValArGspGlnIleAlaIleArgTYrLysAnsGln..... 320
  200 GGTCACCTCATGCTCAGTGCACCTGAA.....AAATCCAGCTCTCCAGT 243
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
  321 ..LeuPhcGlnuAlaThrAspTYrValLysThrThrLysProG1 335
  244 TTTCATATGACAGAAAAAGAGAGAGAAATATGCCAAATATGTC 293
  :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
  335 uTYrTYrSerSerAspLysArgGluuArgGlnAspTYrTYrGln..... 349
  294 CAGTGAATTTGTCAGTGGTGAATTATGCTGATTCCTTTGTTCTTTTA 343
  :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
  350 .....LeuGluAspSer..... 353
  344 TATTTCCAAATTTTATAGACAGACAGAAATACTTTAAGATTTCTGAC 393
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
  354 TYrLeuAlaIlePhcGlnAnsnglyLysTYrLeuGluLysAlaPhcGluuAs 370

```

```

394 TACAAAAAATTGGTTT.....TTTAATCCCTGTGCCAAGTTCAGAAAT 437
  : ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:
  370 nLeuLysProLeuLeuProPhcSerGluLeuGlnInTYrAsnGluLysP 387
  438 ATGCCATA.....ATGGAAACAAGATTAACATAAAGAAG 469
  :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
  387 heTYrLeuGluTYrTYrTrpGlnHisGlyGluAlaArgAspLysSerAsnSP 403
  470 AAAACTACCTCAAGGTTAAATAAACCAAAAAAATAAAAAAATAAAAA 519
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
  404 LysAsnThrThrSerAlaLysLysLysGlnGlnGlnLysLysLysLys 420
  520 GCGGGGCGGAACAGAGGGCGCAA 541
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
  420 sGlyLeuIleLeuArGAsnLys 427

seq_name: pir2:S63322

seq_documentation_block:
  probable membrane protein YNL336w - yeast (Saccharomyces cerevisiae)
  N:Alternate names: hypothetical protein N0275
  C:Species: Saccharomyces cerevisiae
  C>Date: 27-Apr-1996 #sequence_revision 03-May-1996 #text_change 20-Jun-2000
  C:Accession: S63322
  R:Obermayer, B.; Piravandi, E.; Rinke, M.
  submitted to the Protein Sequence Database, April 1996
  A:Reference number: S63317
  A:Accession: S63322
  A:Molecule type: DNA
  A:Residues: 1-381 <OBE>
  A:Cross-references: EMBL:Z71612; NID:g1302462; PIDN:CAA96270.1; PID:g1302463; MIPS:YN
  A:Experimental source: strain S288C
  C:Genetics:
  A:Gene: SGD:COX1
  A:Cross-references: SGD:S0005280; MIPS:YNL336w
  A:Map position: 14L
  C:Superfamily: conserved hypothetical protein YKL219w
  C:Keywords: transmembrane protein
  F:74-90/domain: transmembrane #status predicted <TM1>
  F:236-252/domain: transmembrane #status predicted <TM2>
  F:256-272/domain: transmembrane #status predicted <TM3>

alignment_scores:
  Quality: 77.50      Length: 128
  Ratio: 1.123      Gaps: 6
  Percent Similarity: 53.906      Percent Identity: 24.219

alignment_block:
  US-09-823-101-1 x S63322 ..
  Align seg 1/1 to: S63322 from: 1 to: 381

164 ATGACCTTAATATGAAAAAGAGAGAAATTCCTGTGTAAGTTCATGCT 213
  :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
  59 LeuSerValTrpTrpLysLeuSerAsnAsnCysIleTYrProLeuIleVa 75
  214 CAGTGCACCTGAAAAATCCAGCTCTCCAGTTTTCAT..... 250
  :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
  75 IserLeuLeuValLeuPhcGlnGlyProIlePhcValLeuValIleCysG 92
  251 ..ATGCAGAGAAAAAGAGACAGAGAGAA.....TATGCAGAA 286
  :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
  92 IyLeuSerArgLysArgSerLeuSerLysGlnLeuIleGlnPhcCysLys 108
  287 ANAGTCACAGATATTTG.....TCTGAGTGGTGAATTAAT 321
  :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
  109 GluValThrGluAnsThrProSerSerAspProHisAspTrpGluValVa 125
  322 GCGTGAATTCCTTTGTTTATATTTCCAAATTTTATAGAGACAGAGA 371
  :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
  125 IAlaAlaAsnLeu.....AsnSerTYrL 133

```


THIS PAGE BLANK (USPTO)

OM of: US-09-823-101-1 to: SwissProt_39.* out_format: pfs

Date: Jan 17, 2002 4:08 PM

About: Results were produced by the Gencore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlp
-O=/cgn2.1/USPTO.spool/US09823101/runat.17012002.155138.22061/app-query.fasta.1.2239
-DB=SwissProt_39 -OPMT=fastan -SUFFIX=n2p.rsp -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPEL=0.000 -LOOPEXT=0.000
-OGAPOP=4.500 -OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blonum62
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -MINLEN=0 -MAXLEN=200000000
-USRR=US09823101@cgn1_1_0 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPX
-WAIT -THREADS=1

Search information block:

Query: US-09-823-101-1
Query length: 542
Database: SwissProt_39.*
Database sequences: 100059
Database length: 3664827
Search time (sec): 66.470000

score_list:

Sequence	Strd Orig	ZScore	EScore	Len	Documentation
SwissProt_39:PEX2_HUMAN	89.00	162.30	0.0750	305	p28328 homo sapiens (human)
SwissProt_39:YH16_NEPOL	82.00	150.26	0.4336	247	09tc4 nephroselmis olivacea
SwissProt_39:PEX2_CRIGR	80.00	144.64	0.7251	304	006438 crioctulus griseus (c
SwissProt_39:COS2_YEAST	79.50	141.85	0.8312	379	p38363 saccharomyces cerevis
SwissProt_39:YH1R_YEAST	78.50	137.54	1.08	505	p43132 saccharomyces cerevis
SwissProt_39:COS1_YEAST	78.50	137.88	1.38	381	p53822 saccharomyces cerevis
SwissProt_39:COS7_YEAST	76.50	135.87	1.77	383	007788 saccharomyces cerevis
SwissProt_39:YH13_UREPA	76.00	137.80	1.98	268	09p113 ureplasma parvum (u
SwissProt_39:HST6_CANAL	75.00	122.79	2.74	1532	p53706 candida albicans (ye
SwissProt_39:COS5_YEAST	74.50	131.94	2.93	383	p47187 saccharomyces cerevis
SwissProt_39:PEX2_MOUSE	74.00	132.81	3.29	305	p55098 mus musculus (mouse)
SwissProt_39:HUNK_HUMAN	74.00	125.86	3.43	714	p57058 homo sapiens (human)
SwissProt_39:PEX2_RAT	73.00	130.85	4.24	305	p24392 rattus norvegicus (rat)
SwissProt_39:HOPC_HAEN	73.00	128.51	4.29	406	p44651 haemophilus influenza
SwissProt_39:MOT5_HUMAN	72.50	126.04	4.91	487	015374 homo sapiens (human)
SwissProt_39:NCOM_BUCAI	72.00	124.76	5.58	505	p57263 buchnera aphidicola
SwissProt_39:YH10_YEAST	71.50	136.16	5.89	111	p38834 saccharomyces cerevis
SwissProt_39:YH1R_YEAST	71.50	128.66	6.16	278	p03627 bacteriophage pf3. 32
SwissProt_39:G8R2_MOUSE	71.50	127.84	6.19	307	09e950 mus musculus (mouse)
SwissProt_39:YH10_YEAST	71.50	126.08	6.25	381	p38723 saccharomyces cerevis
SwissProt_39:YH1D_PSEBU	71.00	137.75	6.58	81	p25755 pseudomonas putida. by
SwissProt_39:YH04_PARE	71.00	132.39	6.79	156	p15605 paramecium tetraureli
SwissProt_39:YH04_XENIA	70.00	120.53	9.26	524	004870 xenopus laevis (afric
SwissProt_39:YH04_XENIA	69.50	131.72	9.79	118	034522 fasciola hepatica (l
SwissProt_39:YH04_XENIA	69.50	122.15	10.35	381	p53344 saccharomyces cerevis
SwissProt_39:YH10_HUMAN	69.50	115.29	10.77	882	013342 homo sapiens (human)
SwissProt_39:YH10_HUMAN	68.50	113.06	13.41	437	p26641 homo sapiens (human)
SwissProt_39:YH10_HUMAN	68.50	111.30	13.54	542	009726 schistosoma japonicum
SwissProt_39:YH10_HUMAN	68.50	116.81	13.58	576	p43250 homo sapiens (human)
SwissProt_39:YH10_HUMAN	68.50	114.15	13.80	797	005131 neisseria gonorrhoeae
SwissProt_39:YH10_HUMAN	68.50	114.14	13.80	798	005134 neisseria meningitidis
SwissProt_39:YH10_HUMAN	68.50	111.83	13.82	829	p46954 saccharomyces cerevis
SwissProt_39:YH10_HUMAN	68.00	118.66	15.15	407	p49707 gallus gallus (chicken)
SwissProt_39:YH10_HUMAN	68.00	113.16	15.65	798	087626 neisseria flavescens
SwissProt_39:YH10_HUMAN	68.00	110.92	15.86	1049	p25371 saccharomyces cerevis
SwissProt_39:YH10_HUMAN	67.50	114.20	17.54	623	p46032 candida albicans (yea
SwissProt_39:YH10_HUMAN	67.50	107.67	18.23	1386	p06274 marchantia polymorpha
SwissProt_39:YH10_HUMAN	67.00	117.04	19.46	390	p57558 buchnera aphidicola
SwissProt_39:YH10_HUMAN	67.00	115.95	19.59	446	p54070 saccharomyces cerevis
SwissProt_39:YH10_HUMAN	67.00	115.73	19.61	458	052028 pseudomonas pseudoal

SwissProt_39:PEP_DROME + 67.00 112.08 20.04 716 p1073 drosophila melanoga
SwissProt_39:CHSA_EMBENI - 67.00 109.24 20.37 1013 p30584 emericella nidula
SwissProt_39:YH10_YEAST + 66.50 116.29 22.05 379 p43542 saccharomyces cere
SwissProt_39:HUNK_MOUSE + 66.50 111.12 22.72 714 p08866 mus musculus (mous
SwissProt_39:MOTA_CABEL + 66.50 110.78 22.77 744 p02381 caenorhabditis ele

seq_name: SwissProt_39:PEX2_HUMAN

seq_documentation_block:

ID PEX2_HUMAN STANDARD: PRT: 305 AA.
AC P28328:
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PEROXISOME ASSEMBLY FACTOR-1 (PAF-1) (PEROXIN-2) (PEROXISOMAL MEMBRANE
DE PROTEIN 3) (35 KDA PEROXISOMAL MEMBRANE PROTEIN).
GN PMP3 OR PEX2 OR PAF1 OR PMP35 OR PMP3.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver.
RX MEDLINE=92188187; PubMed=1546315;
RA Shimozawa N., Tsukamoto T., Suzuki Y., Oril T., Shirayoshi Y.,
Mori T., Fujiki Y.;
RT "A human gene responsible for Zellweger syndrome that affects
peroxisome assembly";
RL Science 255:1132-1134 (1992).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX Gartner J.;
RT Submitted (FEB-1992) to the EMBL/GenBank/DBJ databases.
RN [3]
RX CH3C4 ZINC FINGER.
RX MEDLINE=93050151; PubMed=1426230;
RA Patareas R., Fletcher M.A.;
RT "Ring finger in the peroxisome assembly factor-1";
RL FEBS Lett. 312:1-2 (1992).
RN [4]
RP VARIANT IRD LYS-55.
RX Shimozawa N., Imamura A., Zhang Z., Suzuki Y., Oril T., Tsukamoto T.,
Osumi T., Fujiki Y., Wenders R.J.A., Beasley G., Kondo N.;
RT "Defective pxi gene products correlate with the protein import,
biochemical abnormalities, and phenotypic heterogeneity in peroxisome
biogenesis disorders";
RL J. Med. Genet. 36:779-781 (1999).
CC - FUNCTION: SOMEWHAT IMPLICATED IN THE BIOGENESIS OF PEROXISOMES.
CC - SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. PEROXISOMAL.
CC - DISEASE: DEFECTS IN PMP3 ARE THE CAUSE OF ZELLWEGER SYNDROME-1
(ZMS-1), AN AUTOSOMAL RECESSIVE DISORDER DUE TO DEFECTIVE IMPORT
MECHANISMS FOR PEROXISOMAL MATRIX ENZYMES.
CC - DISEASE: DEFECTS IN PMP3 ARE THE CAUSE OF INFANTILE REFSUM
DISEASE (IRD). FEATURES INCLUDE EARLY ONSET, MENIAL RETARDATION,
MINOR FACIAL DYSMORPHISM, RETINITIS PIGMENTOSA, SENSORINEURAL
HEARING DEFICIT, HEPATOMEGALY, OSTEOPOROSIS, FAILURE TO THRIVE,
AND HYPOCHOLESTEROLEMIA. THE BIOCHEMICAL ABNORMALITIES INCLUDE
ACCUMULATION OF PHYTANIC ACID, VERY LONG CHAIN FATTY ACIDS
(VLCFA), DI- AND TRIHYDROXYCHOLESTANOIC ACID AND PIPECOLIC ACID.
CC - SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@sib-sib.ch).
DR EMBL: M86652; AAC12785.1; -
DR EMBL: M85038; AAA60141.1; -


```

123 .....Ty 123
373 TTTCTGCTCTTATAAAATTTGGAATAT.....AAA 339
    ||| ||| ||| ||| ||| ||| ||| ||| |||
123 rheiIeIleuProgluIleItrSerpheuHIsphEgluLyL 140
338 AGAAC.....AAGAAATCAGCCATAATTACACACACAGAC 304
    ||| ||| ||| ||| ||| ||| ||| ||| |||
140 ySerItyrPheasnleuGlnleuAlaIargIleSerSerTyIle 156
303 AAAATCAGTGCATATTTGGCATATTTCCCTCGCTCTTCTCTG 254
    ||| ||| ||| ||| ||| ||| ||| ||| |||
157 GlnPheTrhPheGlnIlePheSerTytrPhePhe...ValIeuPheGlnCy 172
253 CATATGAAACTGAGAGCTGATTTTTCAGTGCACACAGC..... 211
    ||| ||| ||| ||| ||| ||| ||| ||| |||
172 sProleuPheTrh.....HisPheSerIeuasnleuAasnleuLeut 186
210 .....ATGAGTGAGCTTACACAGAAATGTTCTGTTTTCATATT 172
    ||| ||| ||| ||| ||| ||| ||| ||| |||
186 hrIleSerPheLeu.ValasnSerArgIlyTyIleTyrPheLeuPhe 202
171 AAGGTATGTCGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 122
    ||| ||| ||| ||| ||| ||| ||| ||| |||
202 u.....IleuAlaIalPheLeuSerProProAspI 213
121 ACCTCTCT...TTTTCTCTTTTCTTATTTGTTCTAGCCCTTCCCTC 75
    ||| ||| ||| ||| ||| ||| ||| ||| |||
213 IeIuSerIlnPheIleuPheIleuSerIleuIleValPheMetTyrgIuLeu 229
74 ACTTAACCTTAGAGTCTGCTGCTGTTGTTACCATTAAT 39
230 Cys.....ValPheTySerCysPheTyAsp 238

```

seq_name: SwissProt_39:PEX2_CRIGR

```

seq_documentation_block:
ID PEX2_CRIGR STANDARD; PRT; 304 AA.
AC 006438;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PEROXISOME ASSEMBLY FACTOR-1 (PAF-1) (PEROXIN-2) (PEROXISOMAL MEMBRANE
  PROTEIN 3).
GN PMP3 OR PEX2 OR PAF1 OR PMP35.
OS Cricetulus griseus (Chinese hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Cricetulus.
OX NCBI_TaxID=10029;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Ovary;
RX MEDLINE=93286102; PubMed=7685346;
RA Thieringer R., Raetz C.R.H.;
RT "Peroxisome-deficient Chinese hamster ovary cells with point
  mutations in peroxisome assembly factor-1.";
RL J. Biol. Chem. 268:12631-12636(1993).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=94309666; PubMed=8035823;
RA Tsukamoto T., Shimozawa N., Fujiki Y.;
RT "Peroxisome assembly factor 1: nonsense mutation in a peroxisome-
  deficient Chinese hamster ovary cell mutant and deletion analysis.";
RL Mol. Cell. Biol. 14:5458-5465(1994).
CC -1- FUNCTION: SOMEWHAT IMPLICATED IN THE BIOGENESIS OF PEROXISOMES.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. PEROXISOMAL.
CC -1- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```

```

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
DR EMBL: Z17720; CAA78929.1; -.
DR EMBL: D30618; BAA06308.1; -.
DR PIR: S31621; S31621.
DR PIR: A45989; A45989.
DR InterPro: IPR001841; ZnF_fing.
DR Pfam: PF00097; zf-C3HC4; 1.
DR SMART: SM00184; RING; 1.
DR PROSITE: PS00518; ZINC_FINGER_C3HC4; 1.
KW Transmembrane; Peroxisome; Zinc-finger.
FT TRANSMEM 139 158 BY SIMILARITY.
FT ZN_FING 194 212 RING-TYPE.
FT FT 243 282 RING-TYPE.
FT MUTAGEN 246 246 C->Y: LOSS OF FUNCTION.
SO SEQUENCE 304 AA; 34794 MW; 84EC5FA613C148BF CRC64;

```

alignment_scores: Quality: 80.00 Length: 171

Ratio: 0.930 Gaps: 8

Percent Similarity: 50.292 Percent Identity: 22.807

alignment_block:

US-09-823-101-1/rev x PEX2_CRIGR ..

Align seg 1/1 to: PEX2_CRIGR from: 1 to: 304

```

478 GGTAGTTTCTCTTCTTATGTAATCTGTTCAATATGACATATTTCTGAA 429
    ||| ||| ||| ||| ||| ||| ||| ||| |||
148 GlyIuLeuIleasnPheLeuIlePheGlnIlySgIyLysPheAlaTrh 164
428 CTTGGCAGAC.....GGAATTAAAAACCAATTTTGTAGACACA 388
    ||| ||| ||| ||| ||| ||| ||| ||| |||
164 rLeuTrhGluArgLeuLeuGlyIleHisSer...ValPheCysLysProG 180
387 AATTCTTAAGATATTTCTGCTCTCTTATATAAAATTTGGAATATGAA 338
    ||| ||| ||| ||| ||| ||| ||| ||| |||
180 IlnasnIleArgGluValAlGlyPheAspTyrlMetAsnArgIuLeu... 194
337 GAACAAAGATCAGCCATTAATTCACACAGCAACAAATCAGTGACATA 288
    ||| ||| ||| ||| ||| ||| ||| ||| |||
195 .....LeuTrhIlgIyPheAlaIuTrh 202
287 TTTTGGCATATTTCTCTCTGCTGCTTTTCTG..... 256
    ||| ||| ||| ||| ||| ||| ||| ||| |||
202 e.LeuIlePheLeuLeuProLeuIleAsnIleGlnIlySphelYalAlYs 218
255 .....TGCATATAGAAACTGAGAGCTGATTTTTCAGTGC 218
    ||| ||| ||| ||| ||| ||| ||| ||| |||
219 IeuSerSerTrpCysIlePheLeuTrhGlyAlaIleSerAspSerAl 235
217 ACTGAGCATGAGTGCACSTACACAGAAATCTCTGTTTTCATATTAGG 168
    ||| ||| ||| ||| ||| ||| ||| ||| |||
235 AleuAlaIleSerGly.....LysGlnCys..... 243
167 TCATGTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 118
    ||| ||| ||| ||| ||| ||| ||| ||| |||
244 .....AlaIeuCysGlyGluTrpTrhMetProIleTrhIleGly 257
117 CTTCTTTTCTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTAC 68
    ||| ||| ||| ||| ||| ||| ||| ||| |||
258 CysGlnHisValPheCysTytrGlyValIlySerSer.....Phe 272
67 CTTAGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 18
    ||| ||| ||| ||| ||| ||| ||| ||| |||
272 upheAspMetTytrPheTrhCysProIlyCysGlyIleGluValHisSerV 289
    ||| ||| ||| ||| ||| ||| ||| ||| |||
17 TGTGTTCCATG 7
    ||| ||| ||| ||| ||| ||| ||| ||| |||
289 alGlnProLeu 292

```

seq_name: SwissProt_39: COS2_YEAST

```

seq_documentation_block:
ID COS2_YEAST STANDARD: PRT; 379 AA.
AC P38363;
DT 01-OCT-1994 (rel. 30, Created)
DT 01-OCT-1994 (rel. 30, Last sequence update)
DT 20-AUG-2001 (rel. 40, Last annotation update)
DE COS2/COS3 PROTEIN.
OS (COS2 OR YBR302C OR YBR2121) AND (COS3 OR YML132M OR YM4987.03).
OC Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxId=4932;
RN [1]
RP SEQUENCE OF 1-47 FROM N.A. (COS2).
RC STRAIN=S288C;
RA Feldmann H., Mannhaupt G., Schwarzlose C., Vetter I.;
RL Submitted (Aug-1994) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 30-379 FROM N.A. (COS2).
RC STRAIN=S288C;
RA Feuerhann M., Potier S., Souciet J.-L.;
RL Submitted (Aug-1994) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A. (COS3).
RC STRAIN=S288C / AB972;
RA Bowman S., Louis E.J., Barelli B.G., Rajandream M.A., Walsh S.V.;
RL Submitted (Jul-1995) to the EMBL/GenBank/DBJ databases.
CC -1- SUCCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE DUF/COS FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z36171; CAA85267.1; -
DR EMBL: Z50178; CAA90551.1; -
DR PIR: S46187; S46187.
DR SGD: S0000506; COS2.
DR InterPro: IPR001142; DUF.
DR Pfam: PF00674; DUF; 2.
KW Transmembrane; Multigene family.
FT TRANSMEM 73 93 POTENTIAL.
FT TRANSMEM 255 275 POTENTIAL.
FT SEQUENCE 379 AA; 45165 MW; 2AAB51B6C103AF61 CRC64;

```

alignment_scores:
Quality: 79.50 Length: 128
Ratio: 1.187 Gaps: 6
Percent Similarity: 52.344 Percent Identity: 23.438

alignment_block:

US-09-823-101-1 x COS2_YEAST ..

Align seg 1/1 to: COS2_YEAST from: 1 to: 379

```

164 ATGACCTTATATGAAGAAACAGAAATCTGTGTAAGTCGACTCATGCT 213
      ::::::::::: ||||| ::::::::::: ||||| :::::::::::
59 LeuSerValTTrpTyrIleuSerAsnAsnGySIIeYrProleuIleVa 75
214 CAGTGCCACTGAAAAATCCAGCTCTCCAGTTTCTAT..... 250
      :||| ::|||::|||
75 IserLeuValLeuPhelauGlyProIlePhelValLeuValIleGysG 92
251 .. ATGCAGAGAAAAAGACAGAGAGAA.....TATGCCAAA 286

```

```

92 LyleuSerArgLysArgSerIleuSerLysGlnLeuIleGlnPhcGysLys 108
      ::::::::::: ||| :::::::::::
287 ATAGTCACAGTCGATTTG.....TCTGAGTGTGCAATTAT 321
      ::||| ::|||
109 GluIleThrGluAsnThrProSerSerAspProHisAspIrgIuValVa 125
322 GCGTCAATCTTTGTTCTTTTATATTTCCAAATTTTATAGACACAGA 371
      :||| ::|||
125 IAlAlAsnLeuAsnSerIYrLeuYrGluAsnAsnValIrrpAsnThrL 142
372 AATACCTTTAAGAAATTTCTGACTACAAAAAATGCTTTTAAATCCCTG 421
      |||||
142 YsTYrPhe.....PhePhAsnAlaMet 149
422 ...TGCCAA...GTTCAGAAATATGCCATATGACACAGTAACTAAA 465
      |||||
150 ValcysGlnGluAlaIlePheArgThrThrLeuLeuGluIrrProPheSerLeuLys 166
466 GAAGAAACACTACCTCAAGGTTAAAAAACCAAA 499
      ||| ::::::::::: |||
166 slyAspLysAlaIleAlaLysValLysSerPheLys 177

```

seq_name: SwissProt_39: YLIR_YEAST

```

seq_documentation_block:
ID YLIR_YEAST STANDARD: PRT; 505 AA.
AC P43132;
DT 01-NOV-1995 (rel. 32, Created)
DT 01-NOV-1997 (rel. 35, Last sequence update)
DT 01-NOV-1997 (rel. 35, Last annotation update)
DE HYPOTHEICAL 58.3 KDA PROTEIN IN PPRI-SNF7 INTERGENIC REGION (ORFX).
GN YLR015W.
OS Saccharomyces cerevisiae (Baker's yeast)
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxId=4932;
RN [1]
RP SEQUENCE FROM N.A.
RA Vandenbol M., Portetelle D., Hliger F.;
RL Submitted (May-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 1-472 FROM N.A.
RC STRAIN=S288C;
RX MEDLINE=95179524; PubMed=7874498;
RA Levin D.E., Stevenson W.D., Matarabe M.;
RT "Evidence against the existence of the purported Saccharomyces
RT cerevisiae PKC2 gene.";
RL Curr. Biol. 4:990-995(1994).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z73187; CAA97537.1; -
DR EMBL: L34405; AAA34835.1; -
DR SGD: S0004005; YLR015W.
DR InterPro: IPR003877; SPRY.
DR SMART: SM00449; SPRY; 1.
KW Hypothetical protein.
FT SEQUENCE 505 AA; 58347 MW; DSB33221E9F10379 CRC64;

```

alignment_scores:
Quality: 78.50 Length: 174
Ratio: 0.882 Gaps: 7
Percent Similarity: 51.149 Percent Identity: 24.138

alignment_block:

```

alignment_block:
Quality: 78.50 Length: 174
Ratio: 0.882 Gaps: 7
Percent Similarity: 51.149 Percent Identity: 24.138

```


RP SEQUENCE FROM N.A.
 RA Rieger M., Mueller-Auer S., Brueckner M., Schaefer M., Wagner G.;
 RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
 CC -1- SIMILARITY: BELONGS TO THE DUP/COS FAMILY.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

 CC EMBL: Z74296; CAA9828.1; -
 DR SGD: S0002407; COS7.
 DR InterPro: IPR001142; DUP.
 DR Pfam: PF00674; DUP; 2.
 DR Transmembrane; Multigene family.
 KW TRANSMEM 43 63 POTENTIAL.
 FT TRANSMEM 73 93 POTENTIAL.
 FT TRANSMEM 233 253 POTENTIAL.
 FT TRANSMEM 255 275 POTENTIAL.
 SQ SEQUENCE 383 AA; 45679 MW; 5DC6695D06A61663 CRC64;

alignment_scores:

Quality: 76.50 Length: 128
 Ratio: 1.109 Gaps: 6
 Percent Similarity: 53.906 Percent Identity: 22.656

alignment_block:

US-09-823-101-1 x COS7_YEAST ..

Align seg 1/1 to: COS7_YEAST from: 1 to: 383

```

164 ATGACCTTAATGAGAAACAGACATCTTGTCAGTCCATCATGCT 213
      ::::: ||||| ::::: ||||| :::::
59 LeuSerValTrpTrpLysSerAsnScylleTyProLeuIleVa 75
214 CAGGCGCACGMAAATCCAGCTCTCCAGTTTCTAT..... 250
      ::||| ::||| ::|||
75 lSerLeuValIleuPheLeuGlyProIlePheValIleCysG 92
251 . ATGCAGAAAAAGACAGAGAGAA.....TATGCCAA 286
      ::::: ||||| ::::: ||||| :::::
92 lyeuSerAlglysArgSerLeuSerlyGlnLeuIleGlnPheCysLys 108
287 ATAGTCACAGTGAATTG.....TCTGAGTGGTGAATTAT 321
      ::||| ::||| ::|||
109 GluIleThrGlnsInThrProSerSerAspProHisAspTrpGluValVa 125
322 GGCTGATCTTGTCTTTTATATTTCAAATTTTATAAGACAGACA 371
      ::||| ::||| ::|||
125 lValAlaAsnLeu.....AsnSerlyrL 133
372 AATACCTTTAAGAAATTCGACTACAAAAATGGTTTTTAAATTCCTG 421
      ::||| ::||| ::||| ::||| ::||| ::|||
133 euTyrglUasnLysAlaTrpAsnThrLysAsnPhePheAsnAlaThr 149
422 ...TGCCAA...GTTCAGAAATATGCCATAAGACAGATACTATAAA 465
      ::||| ::||| ::||| ::||| ::|||
150 AspCysglUuysMetPheArgThrThrValLeuGlnPhePheSerLeuLys 166
466 GAAGAAACTACTCAAGGTTAAAAAACCAAA 499
      ::||| ::||| ::||| ::|||
166 sLysAspLysAlaAlaLysValLysSerPheLys 177

```

seq_name: Swissprot_39:y131_UREPA

seq_documentation_block:

ID Y131_UREPA STANDARD; PRT; 268 AA.
 AC Q9PR13;

DT 20-AUG-2001 (Rel. 40, Created)
 DT 20-AUG-2001 (Rel. 40, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE HYPOTHETICAL PROTEIN U0131.
 GN U0131.
 OS Ureaplasma parvum (Ureaplasma urealyticum biotype 1).
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
 OC Mycoplasmataceae; Ureaplasma.
 OX NCBL_TaxID=134821;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SEROVAR 3;
 RX MEDLINE=20500219; PubMed=11048724;
 RA Glass J.I., Lefkowitz E.J., Glass J.S., Heiner C.R., Chen E.Y.,
 RA Casseil G.H.;
 RT "The complete sequence of the mucosal pathogen Ureaplasma
 RT urealyticum".
 RL Nature 407:757-762(2000).

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

 CC EMBL: AF002114; AF30537.1; -
 DR Hypothetical protein; Transmembrane; Complete proteome.
 KW TRANSMEM 27 47 POTENTIAL.
 FT TRANSMEM 202 222 POTENTIAL.
 FT TRANSMEM 268 AA; 30686 MW; 01BDC28B5C9A4380 CRC64;
 SQ SEQUENCE

alignment_scores:

Quality: 76.00 Length: 118
 Ratio: 1.246 Gaps: 6
 Percent Similarity: 51.695 Percent Identity: 26.271

alignment_block:

US-09-823-101-1 x Y131_UREPA ..

Align seg 1/1 to: Y131_UREPA from: 1 to: 268

```

80 GAAGGCTAGACAAAATAGGAAAAAGAAAAAGAGGTGCTATTAGTG 129
      ||||| ::||| ::||| ::|||
50 GluIlePheGlnsAlaArgAlaLeuValAsnAspGlnIleIleI 66
130 GGGAAAGGTAGGCCAGACAGCAGCAGACACATGACCTTAATATG. 178
      ||||| ::||| ::||| ::|||
66 eGlyLysValAlaAspValSerGlyAspAsnHisSerLeuMetSerTrpL 83
179 .....AAAAAGAACAT 190
      ||||| ::|||
83 euGlyTyraSerGlnHisAlaAsnTyrglnGluLeuTyrglnIleHis 99
191 TCTGTGTGA.....GGTCCAGTCATG.....CTGCGGCACTGA 225
      ::||| ::||| ::||| ::|||
100 AspGlnLeuLeuAsnGlyLysThrMetGlnSerLeuLeuIleAsnSnThrAs 116
226 AAATCCAGCTCTCCAGTTTCTAT...ATGCAGAAAAAGAGAGAA 272
      ::||| ::||| ::||| ::|||
116 nLysAsnAsnAspProLysLeuTyrgluAlaLeuHisTyHisAlaHis 133
133 eLysPheIleAsnHisAsnGlyThrTrpGlyPheValThrLysValLeu 149
273 GGAATATGCCAA.....ATAGTCACAGTGAATTG 304
      ::||| ::||| ::||| ::|||
305 TCTGAGTGGTGAATATAGGCTGATCTTGTCTTTTATATTCCAAT 354
      ::||| ::||| ::||| ::|||
150 SerAspTrp.....GluAsnSerTrpPheTyrglyTyrglnLe 163
355 TTTT 358

```



```

387 AATCTTAAAGTATTTCTGTGCTGTATTAATAATTTGGAATATATAAA 338
    :: ::::: ::::: ||| ||| :::::
181 llnsmetarglvalglrphcglutyrmetasnargluleu..... 195
337 GAACAAGATGACCATTAATTCACCACTCAGACAAATCTACTGACATA 288
    ||| ::::: :::::
196 .....Leutrhnsiglyrphcalgluph 203
287 TTTTGCAATATTTCTGCTGCTGCTTTTCTC..... 256
    : ||| ::::: ||| :::::
203 e-LeuilepheleuProleuIleasnIleGlLysleuLysAlaLys 219
255 .....TGCATATAGAAAACGTGAGAGCTGATTTTTCAGTGC 218
    ||| ::::: ||| :::::
220 LeuserSerTrpCysThrLeucysThrGlYAlaAlaGlYHnsAspSerTh 236
217 ACTGAGCATGAGTGCACCTACACAAAGATGTTCTGTTCCATATTAAG 168
    ::::: ||| :::::
236 rleuglyserSergly.....Lysglucys..... 244
167 TCATGTGCTGTCTGCTGCTGCTGCTGCTTCCCACTAATGACCT 118
    ::::: ||| ::::: ||| ::::: ||| :::::
245 .....AlaLeucysglYglutPrProthrmProhIsThrIleGly 258
117 CTCCTTTTCTCTTTTCTCTATTTGTTCTAGCCCTCCCTCCTACCTAC 68
    ::::: ||| ::::: ||| :::::
259 CysglHnIsvalrPhcysTrpTrpCysValLysSerSer.....PheLe 273
67 CTTAGAGAGTCTGCTGCTGTTACCAATATGATGAGCTTCACTACAGATC 18
    ::::: ||| ::::: ||| :::::
273 upheAspIleTrpPheThnysProLysCysglYthrGlYvalHnsServ 290
17 TGGTTCCAATG 7
    ::::: ||| :::::
290 alGlnProleu 293

seq_name: SwissProt_39:HUNK_HUMAN

seq_documentation_block:
ID HUNK_HUMAN STANDARD; PRT; 714 AA.
AC P57058;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HORMONALLY UPREGULATED NEU TUMOR-ASSOCIATED KINASE (EC 2.7.1.1-)
DE (SERINE/THREONINE PROTEIN KINASE MAK-V) (BI9).
GN HUNK OR MAKV.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA SCOT H.S., Antonarakis S.E., Papasavvas M.P., Michaud J.;
RT "A putative serine/threonine protein kinase MAK-V on human chromosome
RT 21q22.1".
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
CC -1 SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC SNF1 SUBFAMILY.
CC .....
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC .....
DR EMBL: AJ271722; CAB7146.1; -
DR InterPro: IPR000719; Euk_Pkinase.
DR InterPro: IPR002290; Ser_thr_kin_actsite.
DR Pfam: PF00069; pkinase; 1.

```

```

DR SMART: SM00220; S_TKc; 1.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE: PS00108; PROTEIN_KINASE_ST; 1.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
KW Transferase; Serine/threonine-protein kinase; ATP-binding.
FT DOMAIN 62 320 PROTEIN KINASE.
FT NP_BIND 68 76 ATP (BY SIMILARITY).
FT BINDING 91 91 ATP (BY SIMILARITY).
FT ACT_SITE 186 186 BY SIMILARITY.
SQ SEQUENCE 714 AA; 79685 MW; 87FC85C67DB38AA7 CRC64;

alignment_scores:
    Quality: 74.00 Length: 121
    Ratio: 1.156 Gaps: 5
Percent Similarity: 52.893 Percent Identity: 23.967

alignment_block:
US-09-823-101-1 x HUNK_HUMAN ..

Align seg 1/1 to: HUNK_HUMAN from: 1 to: 714

203 CCACATCATGCTCAGTGCACCTGAAATCCAGCTTCCAGTTTCTATAT 252
    ||| ::::: ||| ::::: ||| :::::
335 ProAsnArgIleSerLeuGlusPleuSerProSerValValLeuHnsIse 351
253 GCAGAGAAAGAGACAGAGAGAAATATGCCAAATATGACACAGATTT 302
    | ::::: ||| ::::: ||| :::::
351 thrGlulys..LeuGlYTrpYAsnSerAspValIleasnThrVal 367
303 TGCTGAGTGTGTAATTTAGCTGATCTTGTCTTTTATTTATTCGA 352
    ||| ::::: ||| ::::: ||| :::::
367 euserAsn.....ArgAlaCysHsIleLeuAlaIleTrpPheLeu 380
353 ATTTTATTAAGACGACAGAAATACTTTTAAGATTTCTGACTACAAA 402
    ::::: ||| ::::: ||| :::::
381 LeuAsnLysLysLeuGlulargTrp.....LeuSerGlYLysSe 393
403 TTGCTTTTAAATTCCTGCTGC.....CAAGTTGAGA 434
    : ::::: ||| ::::: ||| :::::
393 rAspIleGlnsPserLeuCysTrpLysThrArgLeuTrpGlnIleGlu 410
435 AATATGCCATTAATGAGACA.....GATAACTTAAAG 466
    ||| ::::: ||| :::::
410 ySTYArgAlaProLysgluSerTrpGlAlaSerLeuAspThrTrpThr 426
467 AAGAAACTACTCTCAAGTTAAAAAACCAAAAAAATAAAAAA 516
    ::::: ||| ::::: ||| :::::
427 ArgAspLeuGluphenIleAlaValGlnsPlysProLysGluGlnGl 443
517 AAGGGGGGGGGA 529
443 ulYsArgGlyAsp 447

seq_name: SwissProt_39:PEX2_RAT

seq_documentation_block:
ID PEX2_RAT STANDARD; PRT; 305 AA.
AC P24392; O63733;
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PROXISOME ASSEMBLY FACTOR-1 (PAF-1) (PEROXIN-2) (PEROXISOMAL MEMBRANE
DE PROTEIN 3).
GN PAMP3 OR PEX2 OR PAF1 OR PMP35.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY;
RX MEDLINE=91163637; PubMed=1750930;

```



```
301 AATCACTGTGACTAATTTGGCATAATTCCTTCCTTCTTTCTCTGCA 252  
354 ..... SerGly 355  
251 TATAGAAACTGGAGAGCTGGATTTTCAGTGGACATGACATGATGCA 202  
356 TrpValAlaAspGlnAsnTrpIleLysLysTyrHis ..... 367  
201 CCTACACAGAAATGTTCTGTTTCCATATTAAGTCATGTGTCCTGCC 152  
368 . TyrHisLysSerTyrLeuIle ..... LeuCysG 377  
151 TGCCTCTGTGGCCTACCTTCCCACTAATGACCTCTCTTTTCTCTTT 102  
377 LylleThrAsnLeu.LeuAlaProleuAlaThrThr ..... Ph 389  
101 TCCTATTTTG 92  
389 eProleuLeu 392
```



```

439  TGCCTAATGTAAGAACGATACCTAAAGACAAACTACCTCAAGTTAA 488
      ::::: ::::: ||| ||||| ||||| |||
371  ....ValAspAspProGlnAspValPheLysProIhrThrSerArgLeuL 386
      ::::: ::::: ||| ||||| ||||| |||
489  AAAAACCA 496
      ||::: |||
386  YGslnPro 388

seq_name: sp_bacteria:O9JVL6

```

seq_documentation_block:	PRF:	168 AA.
ID_09JVL6	PRELIMINARY:	
AC_09JVL6:		
DT_01-OCT-2000 (TReMBLrel_15, Created)		
DT_01-OCT-2000 (TReMBLrel_15, Last sequence update)		
DT_01-OCT-2000 (TReMBLrel_15, Last annotation update)		
DE_POSSIBLE MEMBRANE PROTEIN:		
CN_NMAN0786:		
OS_Neisseria meningitidis (serogroup A).		
OC_Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria		
OX_NCBI_taxonomy:65699:		

RP SEQUENCE FROM N.A.
RC STRAIN-22491 / SEROGROUP A / SEROTYPE 4A;
RX MEDLINE=20222555; PubMed=10761919;
RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,
RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
RA Davies R.M., Davis P., Devlin K., Feltham T., Hamlin N., Holroyd S.,
RA Jagels K., Leather S., Moule S., Mungall K., Quail M.A.,
RA Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,
RA Thornthwaite S., Spratt B.G., Barrall B.G.;
RT "Complete DNA sequence of a serogroup A strain of *Neisseria*
RT meningitidis 22491";
RL Nature 404:502-506(2000).
DR EMBL; AL162754; CAB84069.1; -.
KW Complete proteome.
SQ SEQUENCE 168 AA; 20297 MW; 91A6G1EE17733412 CRC64;

```

alignment_scores:      length: 174
                       quality: 86.50
                       ratio: 1.095
                       gaps: 9
Percent Similarity:    45.402    Percent Identity: 25.287

alignment_block:
US-09-823-101-1/rev x Q9JVL6    ..

Align seg 1/1 to: Q9JVL6 from: 1 to: 168

```

```

516 TTTTITTTTTTTTTTTTTGGTTTTT.....487
      ||||| ||||| |||||
3 PhePheasnHephegIuNgIyTyrPheIysLysLeuAlaValGluAs 19
486 .....AACCTGGAGGTAGTTCCTCTCTTGTAGTATCTGTTCATTA 444
      ||| ||| ::||| ::||:||||:
19 nGlyLeuAsnThrGluTyrPheIlePheGlnPheIleValCysAsnLeuP 36
443 TGGCATATTTCGAACTTGGCACAGGAAATTAACCAATTTTTCGA 394
      :::: ||| ::::|
36 heSerIlePhe.....ValIleIle 42
393 GTTCGAATAATCTTAATAAGTATTTCTGCTCTTATTAATAAT.....3522
      :: ||| ||| :::::||||| ::
43 LeuMetAsnHepheIlysAsnTyrAlaValLeuIleValIleValI 59
351 .....TGGAAATATTAATAACACACAAGCATATAGCCATTA 318
      ||||| ||| :::::|
59 eValSerIleLeuPheGlyLeuIleLeuIleValAsnLysThrMetSerLeuI 76
317 TTCACACACTGAGCAAA.....ATCACTGTGACT 289
      || :::::|||| ::::|
76 IeLysAsnLysGluLysPheLeuCysGlyLeuPheProLeuAspValPro 92

```

```

288 ATTGTTGCATATTCCCTCGTCCCTTTTCTGCAATAGAAAACGTG 239
      |||      |||      |||      |||      |||      |||
93  ILePhelSYrThmePro.SeTYrPhelYLeuIleSmYsNm 109
238 AGAGCTGATTTTCAGTGCGACCTGAGCATGAGTGCACACAGAAGT 189
      |||      |||      |||      |||      |||      |||
109 eTarGTripLeuphe.....HisMeLan 116
188 GTT.....CTGTTTCATATTAAGTCATGTCGTCTCCCGCTCTC 145
      |||      |||      |||      |||      |||      |||
117 ValPheCYsLeuPhePheIleCYsPheSerIleTYrLeu.....Ileasny 132
144 TGGCCACTTATCCCATCAATAGCACTCTCTTTTCTTTTCCATTT 95
      |||      |||      |||      |||      |||      |||
132 selY.....GuasngInTYrIlealaTYrPheSerCYsTYrPheP 146
94 TTGTCTCAAGCCCTTCCCTC 75
      ||      |||      |||      |||      |||
146 heGlyLeuThrTYrProLeu 152

```

seq_name: sp_virus:089906

seq_documentation_block:	PRELIMINARY;	PRT; 3070 AA.
ID 088906		
AC 088906;		
DT 01-NOV-1998 (Tremblrel. 08, Created)		
DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)		
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)		
DE FUSION PROTEIN.		
OS Beet yellows virus.		
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Closteroviridae;		
OC Closterovirus.		
OX NCBI_TaxID=12161;		
OX 11		
RN SEQUENCE FROM N.A.		
RP STRAIN=CALIFORNIAN.		
RC MEDLINE=98285746; PubMed=9621048;		
RX Peremyslov V.V., Hagihara Y., Dolja V.V.;		
RA "Genes required for replication of the 15.5-kilobase RNA genome of a		
RT plant closterovirus.";		
RL J. Virol. 72:5870-5876(1998).		
NN [2]		

RC	STRAIN=CALIFORMYAN:		
RA	Dolja V., Peremyaslov V., Hagivara Y.;		
RL	Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AF056575; AAC25115.1; -		
DR	InterPro: IPR000606; Viral_helicase1.		
DR	InterPro: IPR001788; RNA_dep.RNAPol2.		
DR	InterPro: IPR002588; V_methyltransf.		
DR	Pfam; PF00978; RNA_dep.RNAPol2; 1.		
DR	Pfam; PF01445; Viral_helicase1; 1.		
DR	Pfam; PF01660; Vmethyltransf; 1.		
FT	CHAIN	1	588
FT	CHAIN	589	2042
FT	PAPAIN-LIKE LEADER PEPTIDE.		
FT	PUTATIVE METHYLTRANSFERASE AND RNA		
FT	HELICASE.		
FT	CHAIN	2626	3070
FT	RNA-DEPENDENT RNA POLYMERASE.		
Q0	SEQUENCE	3070 AA: 345803 MW: 6FC307093FIDF36E	CRC64;

```
alignment_scores:
  quality: 83.50      length: 147
  ratio: 1.14         gaps: 7
Percent Similarity: 49.660  Percent Identity: 25.170
```

```

alignment_block:
US-09-823-101-1/rev x 089906    ..

Align seg 1/1 to: 089906 from: 1 to: 3070

```

448 CATTATGGCAGATATTTCTGAACTTGGCAGCAGGAAATTAAAAACCAATTT 399
||||: :: |||:||||
44 HSPleatqgleThLleSerLeuaspheSerValProaIgdProluPh 60

```

398 TTGGATGACGAATTCTTAAAGATTTTCGTCTCTTAATAAATTGG 349
    ||| :|||
60 eleusnmlairyal..... 65
348 AATATATAAAGAACAAGATCAGCATTAATCCACTACAGACAAAAT 299
    ||| ||| ||| :|||
66 ..... HisleuarigAlaserthrirsapnProleu..... 75
298 CACTGTGACTATTTGGCAATATTTCCCTTCCTGTGCCTTTTCTCGCATAT 249
    ||||| ||| :|||
76 ..... Proserleuproleucilyphehisal 84
248 AGAAACGTC.....AGACGCGATTTTTCAGAG 220
    ||| :|||
84 agluthrphervalleucluleuansnglySerSerAlaproheSerIlep 101
219 GCACGTGACGAT...GACGTGACCACACAGAATGTTTC...TGTTTTCCA 176
    :||| | :||| ||| |||||
101 roserArghHisIleaspPhevalValasnargrproheserValalphero 117
175 TATTAAAGTCATGTGTCTGCCCTGCTCTGTGCGCCACTTTGCCCACT 126
    :||| :||| ||| ||
118 .....ThrgluValleuserValserSerleuarlgthrProserArgle 132
125 AATGACCCCTGTTTCTCTTTCTTTTCCATTTTGTTGCTACCCCTCCGC 76
    ||| :||| :||| :||| :||| :||| :|||
132 uphenIleuencysasp.PphenleutrycserIlysProglyPro 148
75 CACTTAACCTTAGAGTTCGTGCTTGTTACCATTAATGT 37
    ||| :||| ||| |||
149 CysValgluIlealaserPheserThrProProProcys 161

seq_name: sp_virus:O9Q703

seq_documentation_block:
ID O9Q703 PRELIMINARY; PRT: 3090 AA.
AC O9Q703:
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE P348.
OS Beet yellow virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Closteroviridae;
OC Closterovirus.
OX NCBI_TaxID=12161;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN-BYV-4;
RA Peremyslov V.V., Hagiwara Y., Alphanova D., Dolja V.V.;
Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
RL EMBL: AF190581; AAF14300.1; -
DR InterPro: IPR001606; Viral_helicase1.
DR InterPro: IPR001788; RNA_dep_RNAPol12.
DR InterPro: IPR002588; V_methyltransf.
DR Pfam: PF00978; RNA_dep_RNAPol12; 1.
DR Pfam: PF01443; Viral_helicase1; 1.
DR Pfam: PF01660; Vmethyltransf; 1.
SQ SEQUENCE 3090 AA: 347990 MW: 09C779A57DEB74B9 CRC64;
```

[illegible]

```

alignment_block:
US-09-823-101-1/rev x Q9Q703    ..
Align seg 1/1   to: Q9Q703   from: 1   to: 3090

```

```

448 CATATTGCATATTTCGAACTGTGGACAGGAAATTAACCAAAATTT 399
      |||::: :::: ||||::| |::: :::: |||::|
44 HisPheArgLeuThrLeuSerLeuAspPheSerValProArgProLeuPh 60
398 TTGTAGTCAGAATAATCTTAAAGATATTCTGTGCTTTATAAATTTGG 349
      | : : : :
60 eIeuAsnAlaArgVal..... 65
348 AATATATAAAAACAACAGAATCAGCCATAATTCACCAGCTCAGACAAAT 299
      :: ||| ||| :::: ||||| ||
66 .....HisLeuArgAlaSerThrAspAsnProleu..... 75
298 CACTGTGACTATTTTGCGCATTTTCCTTCCTGTCCTTTTCTCGCATAT 249
      ||||| ||::::: |||
76 .....ProSerLeuProLeuGlyPheMetSAl 84
248 AGAAACTGG.....AGAGCTGGAATTTTTCAGTG 220
      |||::: :::: ||| |||
84 aglUthrPheValLeuGluLeuAsnGlySerSerAlaProPheSerIleP 101
219 GCACGTGAGCATT...GAGTGACCTACACAAAGATGTC...TGTTTTCCA 176
      ::::: ||| :||::: :||::: ||| |||||
101 roSeTrArgHIStleAspPheValAlaAsnArgProPheSerValPhePro 117
175 TATTAAAGTCATGNGTCTGCCCTGCTCTGCGGCTACCTTTCCCACT 126
      ::|||::|||::: ||||| ||| ||
118 .....ThglValLeuSerValserSerLeuArgThrProSerArgLe 132
125 AATGACCTCTCTTTTTCTCTTTTCCATTTTGTTCAGCCCTCCCT 76
      |:::: ||| ::| ||||||| |||::| |||::|
132 uPheAlaLeuLeuCySasp.PhePheLeuTyrcysSerLysProGlyPro 148
75 CACTAACCTTAGAGGTTCTGGTCTTGTACCATAAAGT 37
      ::::: :::: ||| |||
149 cysValGluIleValaSerPheSerThrProProCys 161

```

seq_name: sp_invertebrate:Q9N8W5

seq.	annotation block:	
ID	09N8W5	PRELIMINARY; PRT; 99 AA.
AC	09N8W5.	
DT	01-OCT-2000 (TREMBLrel_15	Created)
DT	01-OCT-2000 (TREMBLrel_15	Last sequence update)
DT	01-OCT-2000 (TREMBLrel_15	Last annotation update)
DE	PROBABLE SIMILAR TO RING-H2 FINGER PROTEIN RHA1A.	
GN	CHRI.119.	
OS	Trypanosoma brucei.	
OC	Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma	
NCBI	NCBI_TaxID:5691;	
RP	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN-TREU927;	
RA	Hall N., Bowman S., Quail M., Ivens A.C., Kay M.P., Bray-Allen S.,	
RA	Lennard N.J., Clark L.N., Harris B.R., Melville S., Lawson D.,	
RA	Gerrard C., Rajandream M.A., Barrell B.G.,	
RL	Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.	
DR	EMBL, AL359782; CAB55418.1;	
SO	SEQUENCE 99 AA; 11941 MW; B66DCA5F683CFC064 CRC64;	

alignment_scores:	9
quality:	82.50
length:	32
ratio:	3.438
gaps:	1
percent_similarity:	75.000
percent_identity:	53.125

```
alignment_block:
US-09-823-101-1/rev x Q9N8W5
```



```

156 GGCCTGCTCTGCGCTACCTTTCCCAATAGACCTCTCTTTTCT 107
164 ..... ValThrThrIlePhe 169
106 CTTTTCCTATTGTTGTCAGCCCTCCCTCCTACCTAAC..... 68
170 LLETYAlAlaPheLeuAsnTyPrProPheGlyGluAsnAsnLysAsnAspVa 186
67 .....CTTAGAGTTCTGCTCTGTTCACCAATAGTCAGTCTTCA 28
186 LysTyThrIleuAlaArgPheSerAsnGlyMetGluLeuValAlaIlePhe 203
27 CPGAGATC 20
203 erThrIeu 205

seq_name: sp_bacteria:Q06370

seq_documentation_block:
ID Q06370 PRELIMINARY; PRT; 203 AA.
AC Q06370:
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-MAY-1999 (TReMBLrel. 10, Last annotation update)
DE REPRESSOR PROTEIN OF SURFACE PROTEIN ANTIGEN GENE.
GN PAR.
OS Streptococcus downei (Streptococcus sobrius).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1317;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MT3791;
RX MEDLINE=93322312; PubMed=8331066;
RA Takahashi T., Okanashi N., Hamada S.;
RT "Molecular characterization of a negative regulator of Streptococcus
RT sobrius surface protein antigen gene."
RL J. Bacteriol. 175:4345-4353(1993).
CC -1- FUNCTION: THIS PROTEIN IS A REPRESSOR OF THE SURFACE PROTEIN
CC ANTIGEN GENE.
DR EMBL: D13323; BAA02581.1;
KW Transcription regulation; Repressor; DNA-binding.
SQ SEQUENCE 203 AA; 23395 MW; 9793DB7B40FEFB37 CRC64;

alignment_scores:
Quality: 80.50 Length: 181
Ratio: 1.134 Gaps: 10
Percent Similarity: 39.227 Percent Identity: 25.414

alignment_block:
US-09-823-101-1/rev x Q06370 ..

Align seg 1/1 to: Q06370 from: 1 to: 203

```

```

78 ..... 78
330 GAATGAGCATTAATTACACACTGACAGAAAATACACTGTGACTATTTGGC 281
79 .....HisGlyPheTyrrPheLys 84
280 ATATT...CCTTCCTGTCCTTTTCTGCTCATATTAAGAAACTGAGAGAC 234
85 LeuPheLysProGluAsnProLeuLys...TyrglyThrAspTrpArgLe 100
233 TGGATTTTTCAGTGCAGTGCAGCATGAGTGCAGCTACACAGAAATGTTCT 184
100 uGlyLeuPheAsnAlaThrSerProGlySerArgTyrrProArg..... 114
183 GTTTCATATTAAGTCATGTCGTCTGCTGCTGCTGAGCTACCTT 134
115 .....SerGlnMetLeuIleIleTyrrIeu 122
133 TGCCCA...CTAATGACCTCTCTTTTCTCTTTTCCATTTTGT.... 91
123 AlaProPheValIleuThrSerLeu...LeuLeuThrIleuLeuAlaLeuG 139
90 ....CTTAGCCCTTCCCTCCTTACCTTAACCTTAAGAGTTCTG 55
139 lYThrLeuSerProLeuAlaTyrrIleuPheLeuAlaValIle 152

seq_name: sp_bacteria:Q9KC34

seq_documentation_block:
ID Q9KC34 PRELIMINARY; PRT; 457 AA.
AC Q9KC34:
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TReMBLrel. 15, Last annotation update)
DE BH1740 PROTEIN.
GN BH1740.
OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=86665;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C-125 / JCM 9153;
RX MEDLINE=20512582; PubMed=11058132;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA Fuji F., Hiramata C., Nakamura Y., Ogasawara N., Kuhara S.,
RA Horikoshi K.;
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
RT halodurans and genomic sequence comparison with Bacillus subtilis."
RL Nucleic Acids Res. 28:4317-4331(2000).
DR EMBL: AP001513; BAB05459.1;
KW Complete proteome.
SQ SEQUENCE 457 AA; 51577 MW; 054532562FF36386 CRC64;

alignment_scores:
Quality: 80.50 Length: 195
Ratio: 0.885 Gaps: 9
Percent Similarity: 46.667 Percent Identity: 24.615

alignment_block:
US-09-823-101-1/rev x Q9KC34 ..

Align seg 1/1 to: Q9KC34 from: 1 to: 457

```

```

405 CAATTTTGTAGCAGAAATCTTAAAGTATTTCTGCTCTTATATAA 356
    :|||
92  ..MetPheLeuPheIleProPheIleSerThrPheLeuSerSerValGly 107
355 AATTGGAAATATATAAAGAA.....CAAGAAATCAGCCCAATTAATCCAC 312
    :|||
108 TyrIleAspSerMetLysHisTrpValGlnLysGlnGlnLysArgAr 124
311 ACTGACAGAAATGACGTGTGACTATTTTGGCATATTTCC..... 273
    :|||
124 gLeuArgProAspHisThr..AlaPheTrpLeuThrSerIleValGlyLe 140
272 ..TTCTGTGCTCTT..... 261
    :|||
140 uLeuLeuAsnPheGlySerLeuAlaIleValLysArgIleIleGlyLys 157
260 ..TTCTGTCATATAGAAACTGGAGAGCTGGATTTTTCATGCGACACGA 213
    :|||
157 erPheSerSerPheArgGlu..... 163
212 GCATGAGTGCACCTACACAGATGTTCTGTTTCCATATTAAGTCATG 163
    :|||
164 .....GlnArgLeuMetLeuValIleLeuArgGlyPh 174
162 TGGCTCTGCTGCTCTCTGCTGCTTCCCTTCCACATAAT..... 123
    :|||
174 eGlySerCysLeuLeuTrpSer...ProPheMetValAsnIleGlyLeu 190
122 .....GACCTCTCTTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 87
    :|||
190 leLeuThrIlePheAspLeuSerTrpArgIleGlyGlyLeu 206
86 GCCCTTCCCTCACTTAACCTTAGAGCTCTGTGCTTGTACCATATATGT 37
    :|||
207 leuMetAlaLeuIleTyrIlePheLeuPheTrpLeuPheTrpProAlaI 223
36 GAGCTTCACTCAGATCTGTGTTCCCAATGTCAT 3
    :|||
223 eGlnPheAspHisAspProValIleGlnHisHis 234
seq_name: sp_bacteria:Q9S521

seq_documentation block:
ID   Q9S521          PRELIMINARY;      PRT;       350 AA.
AC   Q9S521.
DT   01-MAY-2000 (TREMBLrel. 13, Created)
DT   01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DE   01-MAY-2000 (TREMBLrel. 13, Last annotation update)
DE   PUTATIVE O ANTIGEN POLYMERASE.
GN   WZY.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN-STOKE W, M92;
RX   MEDLINE=97158136; PubMed=9004408;
RA   Reeves P.R., Hobbs M., Valvano M.A., Skurnik M., Whitfield C.,
RA   Coplin D., Kido N., Klena J., Maskell D., Raetz C.R., Rick P.D.:
RA   "Bacterial polysaccharide synthesis and gene nomenclature.";
RL   Trends Microbiol. 4:495-503(1996).
RN   [3]
RP   SEQUENCE FROM N.A.

```

```

RC   STRAIN-STOKE W, M92;
RX   MEDLINE=98449835; PubMed=9774562;
RA   Wang L., Curod H., Ou W., Reeves P.R.;
RA   "Sequencing of Escherichia coli O111 O-antigen gene cluster and
RT   identification of O111-specific genes.";
RL   J. Clin. Microbiol. 36:3182-3187(1998).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN-STOKE W, M92;
RA   Bastin D.A.;
RL   Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
RN   [5]
RP   SEQUENCE FROM N.A.
RC   STRAIN-STOKE W, M92;
RA   Wang L., Curod H., Ou W., Bastin D.A., Reeves P.R.;
RA   Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
DR   EMBL: AF078736; A046730.1; -
SQ   SEQUENCE 350 AA; 40665 MW; 0E3EA387AC91C1A6 CRC64;

```

```

alignment_scores:
    Quality:      80.00      Length:      229
    Ratio:        0.816      Gaps:        10
    Percent Similarity: 42.795      Percent Identity: 22.707

```

alignment_block:

us-09-823-101-1/rev x Q9S521 ..

Align seg 1/1 to: Q9S521 from: 1 to: 350

```

506 TTTTGTGCTCTTAAATTTTGAACCTGAGCTAGTCTTCT.....TT 463
    ||| ||||| :||| ||||| ||||| ||
98 PheAsnPheLeuTyrLeuLysThrPheAspLysPheSerSerPheProh 114
462 TAGTTATCTGTTCATTAAGCATATTTCTGCACTTGGCACGGCAATT 414
    :||| :||| :||| :||| :||| :|||
114 eThrPhePheIleLeuLeuPheTyrLeuPheArgGlyIleGlyAsnL 131
413 A.....AAAACCAATTTTGTAGTCAGAAATCTTAAAGTAT 373
    |||||
131 euProValTyrLysAsn.....LysLys 138
372 TTCTGTGCTCTTAAATTTTGAATATTAAGAAACAAAGATCAGC 323
    ||| ||||| :||| :||| :|||
139 PheTyrAlaLeuIlePheLeuPheIleLeuIleAspIleMetGlnSerIe 155
322 CATTAATTCACACACTCAGACAAATTCAGCTGACTATTTGGCA..... 280
    :||| :||| :||| :||| :||| :|||
155 uLeuIleAsnTyrArgGlyGlnIleLeuTyrSerValIleCysIleLeu 172
279 .....TATTTTC.....TATTTTC..... 274
    |||||
172 leLeuValPheLysValAsnLeuArgLysIleProTyrPhePheLeu 188
273 ...CTTCTGTGCTCTTAAATTTTGAATATTAAGAA...ACTGAGAGCTGGA 230
    ||||| :||| :||| :|||
189 MetLeuProValLeuTyrValIleIleMetAlaTyrIleGlyPheAsnTyr 205
229 TTTTTCAGTGGCAGCTAGCATGAGCTGAGCTACACAGA..... 190
    :||| :||| :||| :||| :||| :|||
205 rPheAsnLysGlyValThrPhePheGluProThrAlaSerAsnIleGluA 222
189 .....TGTCTGTGTTTC 178
222 rgThrGlyMetIleTyrTyrLeuValSerGlnLeuGlyAspTyrIlePhe 238
177 CAT.....ATTAAGCT 167
    ||| :||| :||| :|||
239 HisGlyMetGlyThrLeuAsnPheLeuAsnAsnGlyGlyGlnTyrLysSh 255
166 CATGTGCTCTGCGCTGCTCTGCGCTACCTTTCCCCAC..... 127
    :||| :||| :||| :||| :||| :|||
255 rLeuTyrGlyLeuProSerLeuIleProAsnAspProHisAspPheLeu 272

```



```

248 AGAAACTGAGAGCTGGATTTCAGTGCACGTGACGATGATGACCT 199
    ::::::::::: |||
40 SerArgLeuSerSerAlaSerPhe..... 47
198 ACACAGACATGTTCTGTTTCCATATTAGTCATGTGGTCTGCGCTGC 149
    |||||::: |||
48 .....LeuPheSerLeuThrLysSer..... 54
148 TCTGTGGCTACCTTTCCCACTAATGACCTCTCTTTTCTCTTTTCC 99
    ||| ::::: ||||| |||||
55 .....GlnProGlnThrGluSerSerPheValPheSer 66
98 TATTTGTTCTAGCCCTTCCC 78
    ||::: ||| |||
67 LeuPheLeuArgAlaAsnPro 73
seq_name: sp.vertibrate:073658

```

```

seq_documentation_block:
ID 073658 PRELIMINARY: PRT: 557 AA.
AC 073658:
DT 01-AUG-1998 (TREMBLrel. 07, Created)
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE OLGRK-C.
OS Oryzias latipes (Medaka fish).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
OC Belontiiformes; Adrianichthyidae; Oryziinae; Oryzias.
OX NCBI_TaxID=8090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98198545; PubMed=9539142;
RA Hisatomi O., Matsuda S., Satoh T., Kotaka S., Imanishi Y.,
RA Tokunaga F.;
RA "A novel subtype of G-protein-coupled receptor kinase, GRK7, in
RT teleost cone photoreceptors.";
RL FEBS Lett. 424:159-164(1998).
CC -1- SIMILARITY: TO THE SER/THR FAMILY OF PROTEIN KINASES. GPRK
CC -1- SIMILARITY: TO THE SER/THR FAMILY OF PROTEIN KINASES. GPRK
CC SUBFAMILY.
DR EMBL: AB009568; BAA25670.1; -.
DR HSSP: P05132; ICRP.
DR InterPro: IPR000719; Euk.pkinase.
DR InterPro: IPR000239; GPCR_kinase.
DR InterPro: IPR000961; Pkinase_C.
DR InterPro: IPR000342; RGS.
DR Pfam: PF00069; pkinase.1.
DR PRINTS: PR00717; GPCR_KINASE.
DR SMART: SM00315; RGS.1.
DR SMART: SM00220; S_TKC.1.
DR SMART: SM00133; S_TK_X.1.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; UNKNOWN_1.
DR PROSITE: PS0011; PROTEIN_KINASE_DOM.1.
DR PROSITE: PS00108; PROTEIN_KINASE_ST.1.
DR PROSITE: PS00132; RGS.1.
KW ATP-binding; Serine/threonine-protein kinase; Transferase.
SQ SEQUENCE 557 AA; 63656 MW; A968753332E0065A CRC64;

```

alignment_scores:

```

Quality: 79.50 Length: 74
Ratio: 2.038 Gaps: 3
Percent Similarity: 52.703 Percent Identity: 33.784

```

alignment_block:

```
us-09-823-101-1 x 073658 ..
```

```
Align seg 1/1 to: 073658 from: 1 to: 557
```

```

344 TATTTCCAAATTTTATAGA.....GC 366
    |||||::: ||:::
166 TyrPheAspLysPheLeuGlnTrpLysGluTrpLysGlnProIle 182
    ::::::::::: |||
367 ACACAAATACCTTTTAAACATTTCTGACTACACAAAAATGGTTTAAAT 416
    ::|||::: ||| |||
182 RasPlySerPheTyrGluPheArgThrLeuGlyLysGlyLysPheGly 199
    ::|||::: |||
417 CCCGTGCGCAAGTTCAGAAATATGCCATTAATGAAACAAGTAACATAAG 466
    ::|||::: |||
199 IuValCysAlaValGln..... 204
467 AGAAACTACCTCAGGTTA.....AAAAACCAAAAAAAA 507
    ||::: ||::: ||:::
205 ValLysAsnTrpGlyGlnMetTyrAlaCysLysLysLeuCysLysLys 221
    ::|||::: |||
508 AAAAAAAAAAAGGGGGCGAA 529
    : |||||::: |||
221 GluLysLysLysGlyGlyGlu 228
seq_name: sp.fungi:004119

```

```

seq_documentation_block:
ID 004119 PRELIMINARY: PRT: 674 AA.
AC 004119:
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE D9461.3/P.
GN PHM5 OR D9461.37 OR YDR452M.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RA Dietrich F.S.;
RA Dietrich F.S.;
RL Submitted (AUG-1995) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Dietrich F.S., Mulligan J., Allen E., Araujo R., Aviles E., Berno A.,
RA Carpenter J., Chen E., Cherry J.M., Chung E., Duncan M.,
RA Hunnicke-Smith S., Hyman R., Komp C., Lashkari D., Lew H., Lin D.,
RA Mosedale D., Nakahara K., Namach A., Oefner P., Oh C., Petel F.X.,
RA Roberts D., Schramm S., Schroeder M., Shogren T., Shroff N.,
RA Winant A., Yelton M., Botstein D., Davis R.W.;
RA Submitted (AUG-1995) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Jia Y., Cherry J.M.;
RL Submitted (JUN-1997) to the EMBL/Genbank/DBJ databases.
DR EMBL: U33027; AAB64872.1; -.
DR SGD: S0002860; PHM5.
SQ SEQUENCE 674 AA; 78344 MW; EEC7B8C7568098B3 CRC64;

```

alignment_scores:

```

Quality: 79.50 Length: 110
Ratio: 1.242 Gaps: 6
Percent Similarity: 58.182 Percent Identity: 29.091

```

alignment_block:

```
us-09-823-101-1 x 004119 ..
```

```
Align seg 1/1 to: 004119 from: 1 to: 674
```

```

256 GAGAAAAAGCAGAGAGAA.....ATATGCCAAATAGTCACAG 296
    |||||::: ||| :::::
463 GluLysAspLysLysLysLysProIleThrArgLysGluLeu 479
    ::|||::: |||
297 TGATTTTGTCTGAGTGTGAATTATGCTGATTCCTTTTATAT 346
    ::|||::: |||
479 egluArgTyrSerIleValAsnIleGlyGlySerValIleProThrPhea 496

```

```

347 TTCCAATTTT.....TATAGACGACGAATACTT..... 378
|||||
496 snProSerPheArgIleTrpGluTyrAsnIleThrAspIleValAsnAsp 512
|||||
379 .....TATAGAAATTTGCTGACTACCAAAAAATTTG.....TTTTTAAT 416
|||||
513 SerAsnIleGluValSerIleTrpGluTyrAsnIleThrAspIleValAsnAsp 529
|||||
417 CCTGTGCAAGTTCAGAAATATGCCATATGCAACAAGAT..... 457
|||||
529 erLeuAsnIleGluValSerIleTrpGluTyrAsnIleThrAspIleValAsnAsp 545
|||||
458 ...AACTAAAGAGAGAAAACTACTCAAGCTTAAAAAACCAAAAAAAA 504
|||||
546 SerAsnIleGluValIleGlyLeuAsnArgGluTyrMetGlyGluTyrLysAs 562
|||||
505 AAAAAAAAAAAAAAAAAAGGGGGGGAACA 532
|||||
562 nLysLysLysLysLysAsnAspLysThr 571

```

seq_name: sp_Invertebrate:Q20109

```

seq_documentation_block:
ID Q20109 PRELIMINARY; PRT; 1398 AA.
AC Q20109;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE F36H1.2 PROTEIN.
GN F36H1.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdilitida; Rhabditoidea;
OC Rhabdilitidae; Peloderinae; Caenorhabditis.
OX NCBL_TaxID=6239;
[1]
RN SEQUENCE FROM N.A.
RP Kershaw J.;
RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Alinscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., McMurtry A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,
RA Smaildon N., Smith A., Sonhammer E., Staden K., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Wellstock L., Wilkinson-Sproat J., Woldman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
elegans."
RL Nature 368:32-38(1994).
DR EMBL; Z68760; CA92996.1; -.
DR HSSP; Q00421; IAWC.
DR InterPro; IPR002110; ANK.
DR InterPro; IPR001412; tRNA-synt_1.
DR Pfam; PF00023; ank; 10.
DR PRINTS; PRO1415; ANKYRIN.
DR SMART; SM00248; ANK; 9.
DR PROSITE; PS00178; AA_tRNA_LIGASE_1; UNKNOWN_1.
DR PROSITE; PS50088; ANK_REPEAT; 7.
DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.
SQ SEQUENCE 1398 AA; 156568 MW; D5080A9E89AB1BEC CRC64;

```

alignment_scores:

Quality:	79.00	Length:	63
Ratio:	2.026	Gaps:	2
Percent Similarity:	61.905	Percent Identity:	33.333

alignment_block:
US-09-823-101-1/rev x Q20109 ..
Align seg 1/1 to: Q20109 from: 1 to: 1398

```

519 TTTTNTTTTTTTTTTTTTTTTTTTTTTTTACCTTGAGGTGTTT 470
|||||
491 PhePheAlaIlePheLeuPhePheGlyMetPheSerLeuTrpPheThrMe 507
|||||
469 CTTCCTTTAGTTATCTGTGTCATATGCAATATTCGAACCTTGACACA 420
|||||
507 tLeuIleAlaIleSerAsnSerValThrAlaTyrIleIleSerTrpSerV 524
|||||
419 GGGAAATTAAAAAACCAATTTTGTAGTCAGAAATTCCTAAAAATATTTC 370
|||||
524 alPheLeu.....LeuIlePheIleIle.....Phe 532

```

```

369 TGTGCTCTTATAAAATTTGGAATATATAAGAAACAA 331
|||||
533 CysSerLeuIleValValValTyrTrpGlyAspArgLys 545

```

```
US-09-823-101-2 x AAE02399 ..
Align seg 1/1 to: AAE02399 from: 1 to: 382

39 CNAAGAGTAGAGAGATAGAGTCAGGCTCTATAGCAAACTACTATCG 88
   ::::::::::::::::::::
311 ArgGlyIylArgIylArgIylArgIyl..... 317
89 CTTAACTCCCGAGGAAGGAGATCCAAACCTGGAAGAACAGACGA 138
   ::::::::::::::::::::
318 ...ArgSerArgGlyArgGlyGly.....GlyArgArgGlyGly 330
139 GTAAAGTGGG...CGAGTGGAGGCGAGGAAGCTTGGTGACAGATCAG 185
   ::::::::::::::::::::
330 eArgAlaGlyArgArgGlyGlyGlyArgArgGly.....Arg 342
186 GGAAGG 191
   |||||
343 GlyGly 344

seq_name: /SID8/gcdata/geneseq/geneseq/AA2000.DAT:AA51860

seq_documentation_block:
ID AAB51860 standard; Protein: 124 AA.
XX
AC AAB51860;
XX
DT 16-FEB-2001 (first entry)
XX
DE Human secreted protein sequence encoded by gene 34 SEQ ID NO:93.
XX
KW Human; secreted protein; immunosuppressive; antiarthritic; antirheumatic;
KW antiproliferative; cytostatic; cardiant; vasotropic; cerebroprotective;
KW neutrophic; neuroprotective; antibacterial; virucide; fungicide;
KW ophthalmological; vulnery; autoimmune disease; rheumatoid arthritis;
KW hyperproliferative disorders; cancer; cardiovascular disorder;
KW cardiac arrest; cerebrovascular disorder; nervous system disorder;
KW Alzheimer's disease; ocular disorder; wound healing; skin aging.
XX
OS Homo sapiens.
XX
PN WO200061626-A1.
XX
PD 19-OCT-2000.
XX
PF 06-APR-2000; 2000WO-US09066.
XX
PR 09-APR-1999; 99US-0128698.
XX
PR 20-JAN-2000; 2000US-0176926.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PA (ROSE/) ROSEN C A.
XX
PI .Rosen CA, Ruben SM, Komatsoulis G;
XX
DR WPI: 2000-619227/59.
XX
DR N-PSDB: AAC93512.
XX
PT New nucleic acid molecules encoding 49 human secreted proteins for
PT diagnosing, preventing or ameliorating medical conditions and used for
PT food additives or preservatives -
XX
XX
XX Claim 11; Page 476; 51pp; English.
XX
XX Polynucleotide sequences AAC93479 - AAC93527 represent cDNA encoding
CC human secreted proteins AAB51827 - AAB51875. Sequences AAB51876 -
CC AAB51927 represent alternative polypeptides encoded by the genes, and
CC amino acid sequences with which they share homology. The genes and
CC proteins have activities dependent on the tissues and cells in which they
CC are expressed. Examples of their activities include immunosuppressive;
CC antirheumatic; antiproliferative; cytostatic; cardiant;
CC vasotropic; cerebroprotective; neutrophic; neuroprotective; antibacterial;
CC virucide; fungicide; ophthalmological; and vulnery. The secreted
```

```
CC proteins, polynucleotides, antagonists and agonists may be useful in
CC treating, preventing and/or diagnosing diseases and disorders such as
CC autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative
CC disorders e.g. neoplasms of the breast or liver, cardiovascular disorders
CC e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia,
CC anglogenesis, nervous system disorders e.g. Alzheimer's disease,
CC infections caused by bacteria, viruses and fungi and ocular disorders
CC e.g. corneal infection. The polypeptides can also be used to aid wound
CC healing and epithelial cell proliferation, to prevent skin aging due to
CC sunburn, to maintain organs before transplantation, for supporting cell
CC culture of primary tissues, to regenerate tissues and in chemotaxis. The
CC polypeptides can also be used as a food additive or preservative to
CC increase or decrease storage capabilities, fat content, lipid, protein,
CC carbohydrate, vitamins, minerals, cofactors and other nutritional
CC components. Oligonucleotides AAC93470 - AAC93478 and peptide AAB51826 are
CC used in the isolation and characterisation of the proteins and
CC polynucleotides of the invention.
XX
SQ Sequence 124 AA;
XX

alignment_scores:
Quality: 74.00 Length: 35
Ratio: 3.083 Gaps: 1
Percent Similarity: 68.571 Percent Identity: 42.857

alignment_block:
US-09-823-101-2/rev x AAB51860 ..
Align seg 1/1 to: AAB51860 from: 1 to: 124

185 CTGATCTGTCACCCACCTTCTCCTCCACCGCCACCTTCTACTCG 136
   ||| |||||::: ||::||| |||||:::
28 LencysHisProAlaProProSerSerSerProThrserse 44
135 CTGTCCTTCTTC.....AGGTTTGATCTCCCTTCCTCGG 98
   ||| ::||| :: ||:::||||:
44 rLeucysAlaserValserCysArgIylsTrpValGluProGluArg 61
97 GACTG 93
   |||
61 rGlu 62

seq_name: /SID8/gcdata/geneseq/geneseq/AA2000.DAT:AA43199

seq_documentation_block:
ID AAB43199 standard; Protein: 526 AA.
XX
AC AAB43199;
XX
DT 08-FEB-2001 (first entry)
XX
DE Human ORFX ORF2963 polypeptide sequence SEQ ID NO:5926.
XX
KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW vulnery; antiporiatic; antiparkinsonian; neutrophic; neuroprotective;
KW anticonvulsant; osteopathic; antirheumatic; immunosuppressant; cardiant;
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW antinaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW thrombosis; contraceptive.
XX
XX Homo sapiens.
XX
XX WO200058473-A2.
XX
XX
```



```

PD      05-OCN-2000.
XX
PF      31-MAR-2000; 2000MO-US08621.
PR
PR      31-MAR-1999; 99US-0127607.
PR      02-APR-1999; 99US-0127636.
PR      05-APR-1999; 99US-0127728.
PR      30-MAR-2000; 2000US-0540763.
PA
PA      (CURA-) CURAGEN CORP.
PI
PI      Shinkets RA, Leach M;
DR
DR      WPI: 2000-602362/57.
XX
XX      N-PSDB; AAC77408.
XX
XX      Novel nucleic acids and peptides derived from open reading frame X,
PT      useful for treating e.g. cancers, proliferative disorders,
PT      neurodegenerative disorders and cardiovascular disease -
XX
PS      Claim 11: Page 5111-5112; 5507pp; English.
XX
XX      AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC      which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC      sequences have activities such as: cytostatic; hepatotropic; vulnery;
CC      antiparasitic; antiparkinsonian; neotrophic; neuroprotective;
CC      osteopathic; anticoagulant; antiarthritic; immunosuppressant;
CC      immunostimulant; cardiact; thrombolytic; coagulant; vasotrophic;
CC      antidiabetic; hypotensive; dermatological; immunosuppressive;
CC      antinflammatory; antibacterial; antiviral; antifungal; antineumatic;
CC      antithyroid; and antianaemic. The sequences can be used for determining
CC      the presence of or predisposition to, or preventing or treating
CC      pathological conditions associated with an ORFX-associated disorder. The
CC      nucleic acids can be used to express ORFX proteins in gene therapy
CC      vectors. The proteins and nucleic acids may be used to treat cancers,
CC      proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC      graft vs host disease, cardiovascular disease, diabetes mellitus,
CC      hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
CC      erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC      bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC      allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC      nocturnal hemoglobinuria, antinflammatory disease; to enhance
CC      coagulation; to inhibit thrombosis; and as a contraceptive.
XX
XX      Sequence 526 AA:
SQ
    Alignment_scores:
        Quality: 73.00          Length: 51
        Ratio: 2.920           Gaps: 3
    Percent Similarity: 49.020   Percent Identity: 43.137
alignment_block:
US-09-823-101-2/rev x AAB43199 ..
Align seg 1/1 to: AAB43199 from: 1 to: 526
190 CTTCCGCTATGTGCACCACTTTCCTGCC...ACTCGCCACT 144
    ||||| ||||| ||||| ||||| |||||
27 ProProLeuSerValProGlnProPheLeuProValPierHMetProLe 43
143 TCTACTGCTTGTCCTCTTCAGGGTTTGATCTCCCTCCTCGGGACT 94
    |||| | || ||||| ||
43 uLeuSerPro.....SerProLaIraProPropoi 53
93 GTAACCCTAGCATTTGGCTATGAGACCTCACTTATCTCTCTACT 44
    ::::||| |||::|||
53 lSerPro.....ValLeuProLeuValPro 61
43 CCT 41
    |||
62 pro 62

```

```

seq_name: /S1058/gcgdata/geneseq/geneseq/AA2000.DAT.AAB42882
seq_documentation_block:
ID   AAB42882 standard; Protein; 142 AA.
XX
XX   AAB42882:
XX
XX   08-FEB-2001 (first entry)
XX
XX   Human OREF. OREF2646 polypeptide sequence SEQ ID NO:5292.
XX
XX   Human: open reading frame; OREF; detection: cytosstatic; hepatotropic;
XX   vulnerrary; antipsoiatric; antiparkinsonian; noctropic; neuroprotective;
XX   anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
XX   immunostimulant; thrombolytic; coagulant; vasotrophic; antidiabetic;
XX   hypotensive; dermatological; immunosuppressive; antinflammatory;
XX   antiviral; antibacterial; antifungal; antirheumatic; antihydroid;
XX   antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
XX   neurodegenerative disorder; osteoarthritis; graft vs host disease;
XX   cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
XX   cholesterol ester storage; systemic lupus erythematosus; infection;
XX   severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
XX   allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
XX   bone damage; cartilage damage; antinflammatory disease; coagulation;
XX   thrombosis; contraceptive.
XX
XX   Homo sapiens.
XX
XX   WO200058473-A2.
XX
XX   05-OCT-2000.
XX
XX   31-MAR-2000; 2000WO-US08621.
XX
XX   31-MAR-1999; 99US-0127607.
XX   02-APR-1999; 99US-0127636.
XX   05-APR-1999; 99US-0127728.
XX   30-MAR-2000; 2000US-0540763.
XX
XX   (CURA-) CURAGEN CORP.
XX
XX   Shilmkets RA, Leach M;
XX
XX   WPI: 2000-602362/57.
XX
XX   N-PSDB: AAC77091.
XX
XX   Novel nucleic acids and peptides derived from open reading frame X,
XX   useful for treating e.g. cancers, proliferative disorders,
XX   neurodegenerative disorders and cardiovascular disease -
XX
XX   Claim 11; Page 4461-4462; 5507pp; English.
XX
XX   AAC77446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
XX   which represent the human OREF open reading frames 1 to 3161. The OREF
XX   sequences have activities such as: cytostatic; hepatotropic; vulnerrary;
XX   antipsoiatric; antiparkinsonian; noctropic; neuroprotective;
XX   osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
XX   immunostimulant; cardiant; thrombolytic; coagulant; vasotrophic;
XX   antidiabetic; hypotensive; dermatological; immunosuppressive;
XX   antinflammatory; antibacterial; antiviral; antifungal; antineumatic;
XX   antihydroid; and antianaemic. The sequences can be used for determining
XX   the presence of or predisposition to, or preventing or treating
XX   pathological conditions associated with an OREF-associated disorder. The
XX   nucleic acids can be used to express OREF proteins in gene therapy
XX   vectors. The proteins and nucleic acids may be used to treat cancers,
XX   proliferative disorders, neurodegenerative disorders, osteoarthritis,
XX   graft vs host disease, cardiovascular disease, diabetes mellitus,
XX   hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
XX   erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
XX   bacterial or fungal infection, malaria, autoimmune disorders, asthma,
XX   allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
XX   nocturnal haemoglobinuria, antinflammatory disease; to enhance
XX   coagulation; to inhibit thrombosis; and as a contraceptive.

```


PA (YAMA) YAMANOUCHI UK LTD.
XX
XX Kellie S, Finan P.
XX
XX WPI; 2000-163495/15.
XX
XX Src homology 3 protein or fragment for preventing or treating
PT proliferative disease, such as cancer or chronic inflammatory disease,
PT comprises src homology 3 domain binding activity and nuclear
PT localization activity.
XX
PS Claim 5; Page 35; 52pp; English.
XX
XX The invention provides a novel protein (np70) that has src homology 3
CC (SH3) domain binding activity and nuclear localisation activity. The
CC proline-rich SH3 binding protein np70 or its fragments are useful for
CC the prevention or treatment of proliferative disease, such as cancer and
CC chronic inflammatory disease. The present sequence represents the
CC np70 protein carboxy terminal region.
XX
SQ Sequence 250 AA;

alignment_scores:
Quality: 68.50 Length: 64
Ratio: 2.635 Gaps: 2
Percent Similarity: 40.625 Percent Identity: 32.812

alignment_block:
US-09-823-101-2/rev x AAY67470 ..

Align seg 1/1 to: AAY67470 from: 1 to: 250

190 CTTCCCTGATCTGTCAACCACTTTCTGCTCCCACT..... 152
|||||
25 ProProLaProProLeuArgProProGlyProProthrlGlyLeuProPr 41
151CGCCCACTTCTACTGCTTGTCTTCTTCAG 121
|||||
41 oglyProProProGlyAlaProProPhleuArgProProGlyMetProG 58
120 GGTTCGATCTCCCTTCCTGCGGACTGTAGCATAGTATTGGCTATG 71
|||||
58 lyeuArgGlyProLeuProArg..... 65
70 AGAGCTGACTTATCTCTACTACTCTTGACACCCCA 29
|||||
66LeuLeuProProGlyProProPro 73

seq_name: /SID58/gcgcdata/geneseq/geneseqp/AA2000.DAT: AAY82327

seq_documentation_block:

ID AAY82327 standard; Protein; 641 AA.

XX AAY82327;

XX 21-JUN-2000 (first entry)

DE Human Npw38BP1 transcription related protein SHQ ID NO:1.

KW Human; Npw38BP1; Npw38; transcription related protein; diagnosis;
KW genetic disease; gene therapy; antitumour.

XX Homo sapiens.

XX JP2000060562-A.

XX 29-FEB-2000.

XX 21-AUG-1998; 98JP-0235901.

XX 21-AUG-1998; 98JP-0235901.

XX

PA (KAGA-) KAGAKU GIJUTSU SHINKO JIGYODAN.
PA (KATO/) KATO S.
XX
XX WPI; 2000-249677/22.
XX
XX N-PSDB; AAA08075.
XX
XX Human transcription-related protein and a cDNA encoding said protein
PT useful for screening for new antitumor agents and for diagnosis
PT
XX
PS Claim 1; Page 6-7; 10pp; Japanese.
XX
XX The present sequence represents a human transcription-related protein.
CC The protein and an antibody against the protein, are useful for
CC diagnosis and treatment of various diseases. The human cDNA is useful
CC as a probe for genetic diagnosis and a gene source for gene therapy.
CC A new antitumor agent can be detected by screening a low molecular
CC compound combining with the protein.
XX
SQ Sequence 641 AA;

alignment_scores:
Quality: 68.50 Length: 64
Ratio: 2.635 Gaps: 2
Percent Similarity: 40.625 Percent Identity: 32.812

alignment_block:
US-09-823-101-2/rev x AAY82327 ..

Align seg 1/1 to: AAY82327 from: 1 to: 641

190 CTTCCCTGATCTGTCAACCACTTTCTGCTCCCACT..... 152
|||||
416 ProProLaProProLeuArgProProGlyProProthrlGlyLeuProPr 432
151CGCCCACTTCTACTGCTTGTCTTCTTCAG 121
|||||
432 oglyProProProGlyAlaProProPhleuArgProProGlyMetProG 449
120 GGTTCGATCTCCCTTCCTGCGGACTGTAGCATAGTATTGGCTATG 71
|||||
449 lyeuArgGlyProLeuProArg..... 456
70 AGAGCTGACTTATCTCTACTACTCTTGACACCCCA 29
|||||
457LeuLeuProProGlyProProPro 464

seq_name: /SID58/gcgcdata/geneseq/geneseqp/AA2000.DAT: AAY67469

seq_documentation_block:

ID AAY67469 standard; protein; 641 AA.

XX AAY67469;

XX 12-MAY-2000 (first entry)

DE Np70 protein sequence.

KW SH3; np70; src homology 3; nuclear localisation; anticancer;
KW anti-proliferative; anti-inflammatory.

XX Homo sapiens.

XX GB2341182-A.

XX 08-MAR-2000.

XX 01-SEP-1998; 98GB-0019038.

XX 01-SEP-1998; 98GB-0019038.

XX (YAMA) YAMANOUCHI UK LTD.

XX

PI Kellie S, Finan P;
 XX
 DR WPI: 2000-163495/15.
 XX
 PT SRC homology 3 protein or fragment for preventing or treating
 PT proliferative disease, such as cancer or chronic inflammatory disease,
 PT comprises src homology 3 domain binding activity and nuclear
 PT localization activity -
 XX
 PS Claim 5; Page 35; 52pp; English.
 XX
 CC The invention provides a novel protein (np70) that has src homology 3
 CC (SH3) domain binding activity and nuclear localisation activity. The
 CC proline-rich SH3 binding protein np70 or its fragments are useful for
 CC the prevention or treatment of proliferative disease, such as cancer and
 CC chronic inflammatory disease. The present sequence represents the
 CC np70 protein.
 CC
 XX
 SQ Sequence 641 AA;
 alignment_scores:
 Quality: 68.50 Length: 64
 Ratio: 2.635 Gaps: 2
 Percent Similarity: 40.625 Percent Identity: 32.812
 Alignment_block:
 US-09-823-101-2/rev x AAY67469 ..
 Align seg 1/1 to: AAY67469 from: 1 to: 641
 190 CCTCCCTGATCTGTACCCACCTTCTGCTCCACT..... 152
 ||||| :||| |||||
 416 ProProAlaProProLeuAlaProProGlyProProThGlyLeuProPr 432
 151CGCCACTTCTACTGCTTGTCTTCCAG 121
 |||:| ||| |||
 432 oGlyProProProGlyAlaProProPhleuAlaProProGlyMetProG 449
 120 GGTTCGATCTCCCTCTCGGACTGTAAACCATAGATTGGCTATG 71
 |||:| :||| |||||
 449 lYleuAlaGlyProleuProAlaG..... 456
 70 AGAGCTGACTCTTATCTCTCTACCTCTTGACACCCCA 29
 ||||| |||||
 457LeuLeuProProGlyProProPro 464
 seq_name: /SIDS8/gcgdata/geneseq/geneseq/AA2001.DAT:AAV72165
 seq_documentation_block:
 ID AAV72165 standard; Protein: 641 AA.
 XX
 AC AAV72165;
 XX
 DT 24-APR-2001 (first entry)
 XX
 DE Human RNA metabolism protein (RMEP-5).
 XX
 KW Human: RNA metabolism protein; RMEP; nootropic; neuroleptic; antilcer;
 KW transquilliser; antianemic; antidiabetic; immunosuppressive; cytostatic;
 KW antiaesthetic; antiinflammatory; anti-HIV; human immunodeficiency virus;
 KW antiarthritic; antiatherosclerotic; antiatherosclerotic; antiallergic;
 KW antirheumatoid; antiparkinsonian; antithyroid; nephrotoxic; antipain;
 KW thyromimetic; RMEP expression modulator; transgenic; spinal cord disease;
 KW nervous system disorder; Alzheimer's disease; therapy; gene therapy;
 KW neuromuscular disorder; hepatitis; cancer; cell proliferative disorder;
 KW peripheral nervous system disorder; cirrhosis; cranial nerve disorder;
 KW progressive neural autonomic nervous system disorder; Addison's disease;
 KW amyotrophic lateral sclerosis; autoimmune disorder; drug screening.
 XX
 XX Homo sapiens.
 OS
 XX
 PN WO200078952-A2.

XX
 PD 28-DEC-2000.
 XX
 PF 15-JUN-2000; 2000WO-US16644.
 XX
 PR 17-JUN-1999; 99US-0139922.
 XX
 PA (INCY-) INCYTE GENOMICS INC.
 XX (BAUG/) BAUGHN M R.
 XX
 PI Bandman O, Yue H, Lal P, Tang YT, Reddy R, Azimzai Y;
 XX
 DR WPI: 2001-102723/11.
 DR N-PSDB; AAD02348.
 XX
 PT New human RNA metabolism proteins (RMEP), useful for diagnosing,
 PT treating, preventing nervous system, cell proliferative,
 PT autoimmune/inflammatory disorders associated with abnormal expression
 PT of RMEP
 XX
 PS Example 1; Page 86-87; 103pp; English.
 XX
 CC The present sequence is human RNA metabolism protein (RMEP-5) encoded by
 CC a cDNA (clone ID 1858421) obtained from PROSNOT18 cDNA library.
 CC Agonists and antagonists of RMEP cDNA are useful for treating diseases or
 CC conditions associated with altered expression of functional RMEP. RMEP
 CC sequence or their mammalian homologues are useful for creating 'knock
 CC out' or 'knock in' humanised animals or transgenic animals to model
 CC human disease. RMEP sequence is useful in the diagnosis, prevention and
 CC treatment of nervous system disorders e.g. Alzheimer's disease, Pick's
 CC disease, Huntington's disease, Parkinson's disease, amyotrophic lateral
 CC sclerosis, and other motor neuron disorders, progressive neural autonomic
 CC nervous system disorders, cranial nerve disorders, spinal cord diseases,
 CC muscular dystrophy and other neuromuscular disorders, peripheral nervous
 CC system disorders, mental disorders including anxiety and schizophrenia,
 CC anorexia etc, cell proliferative disorders e.g. actinic keratosis,
 CC arteriosclerosis, atherosclerosis, cirrhosis, hepatitis, mixed connective
 CC tissue disease (MCTD), cancers e.g. adenocarcinoma, leukemia, lymphoma,
 CC melanoma etc., and autoimmune/inflammatory disorders such as acquired
 CC immuno deficiency syndrome (AIDS), Addison's disease, allergies, anaemia,
 CC asthma, diabetes mellitus, rheumatoid arthritis, Grave's disease and
 CC autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APCED).
 CC RMEP cDNA is useful for somatic or germ-line gene therapy. RMEP sequence
 CC is useful several drug screening assays.
 CC
 XX
 SQ Sequence 641 AA;
 alignment_scores:
 Quality: 68.50 Length: 64
 Ratio: 2.635 Gaps: 2
 Percent Similarity: 40.625 Percent Identity: 32.812
 Alignment_block:
 US-09-823-101-2/rev x AAV72165 ..
 Align seg 1/1 to: AAV72165 from: 1 to: 641
 190 CCTCCCTGATCTGTACCCACCTTCTGCTCCACT..... 152
 ||||| :||| |||||
 416 ProProAlaProProLeuAlaProProGlyProProThGlyLeuProPr 432
 151CGCCACTTCTACTGCTTGTCTTCCAG 121
 |||:| ||| |||
 432 oGlyProProProGlyAlaProProPhleuAlaProProGlyMetProG 449
 120 GGTTCGATCTCCCTCTCGGACTGTAAACCATAGATTGGCTATG 71
 |||:| :||| |||||
 449 lYleuAlaGlyProleuProAlaG..... 456
 70 AGAGCTGACTCTTATCTCTCTACCTCTTGACACCCCA 29
 ||||| |||||
 457LeuLeuProProGlyProProPro 464
 seq_name: /SIDS8/gcgdata/geneseq/geneseq/AA2001.DAT:AAV72165
 seq_documentation_block:
 ID AAV72165 standard; Protein: 641 AA.
 XX
 AC AAV72165;
 XX
 DT 24-APR-2001 (first entry)
 XX
 DE Human RNA metabolism protein (RMEP-5).
 XX
 KW Human: RNA metabolism protein; RMEP; nootropic; neuroleptic; antilcer;
 KW transquilliser; antianemic; antidiabetic; immunosuppressive; cytostatic;
 KW antiaesthetic; antiinflammatory; anti-HIV; human immunodeficiency virus;
 KW antiarthritic; antiatherosclerotic; antiatherosclerotic; antiallergic;
 KW antirheumatoid; antiparkinsonian; antithyroid; nephrotoxic; antipain;
 KW thyromimetic; RMEP expression modulator; transgenic; spinal cord disease;
 KW nervous system disorder; Alzheimer's disease; therapy; gene therapy;
 KW neuromuscular disorder; hepatitis; cancer; cell proliferative disorder;
 KW peripheral nervous system disorder; cirrhosis; cranial nerve disorder;
 KW progressive neural autonomic nervous system disorder; Addison's disease;
 KW amyotrophic lateral sclerosis; autoimmune disorder; drug screening.
 XX
 XX Homo sapiens.
 OS
 XX
 PN WO200078952-A2.

Ratio: 2.833 Gaps: 2
Percent Similarity: 41.379 Percent Identity: 27.586

alignment_block:

US-09-823-101-2/rev x AAM66399 ..

Align seg 1/1 to: AAM66399 from: 1 to: 42

```

199 GTCAGACACCTCCGATCTGTACACCAACCTTCTCCGCTCCACTCG 150
    :::::|||||
6  IleaArgProPro.....IleaArgProPhetProProhear 20
149 CCCACTTCTACTGCTTGTCTTCTCCAGGTTTGATCTCCCTTCCTC 100
    ||| ::| ||| :::::|
20 gProProIleaArgProProIlehearPro..... 29
99 GGGACTGTAAAGCCATAGTATTGGCTATAGAGCTGACTTATCTCTC 50
    ||| ::| ||| :::::|
30 .....ProIle 31
49 CTACTCTCTTGACACCCCACTC 26
    |||||::: |||||
32 ArgProProhearProProleu 39

```

seq_name: /SID88/gcdata/geneseq/geneseq/AA2000.DAT:AAV91698

seq_documentation_block:

ID AAV91698 standard: Peptide: 42 AA.

AC AAV91698;

DT 06-JUN-2000 (first entry)

DE Cationic peptide Bac5 amino acid sequence.

KM Cationic peptide; tumour; pharmaceutical composition; cancer; treatment;
KM leukaemia; polyoxalkylene-modified; APO; lymphoma; multiple myeloma;
KM breast; lung; ovary; cervix; uterus; skin; prostate; liver; colon;
KM multidrug resistance.

OS Unidentified.

PN W09965506-A2.

PD 23-DEC-1999.

PF 14-JUN-1999; 99WO-CA00552.

PR 12-JUN-1998; 98US-0096541.

PA (MICR-) MICROLOGIX BIOTECH INC.

PI Friedland HD, Krieger TJ, Taylor R, Erfle D, Fraser JR, West MHP;

DR WPI; 2000-223549/19.

PT Novel pharmaceutical composition containing optionally activated
PT polyoxalkylene-modified cationic peptides, useful for treating tumours

PS Disclosure; Page 9; 94pp; English.

CC This sequence represents a cationic peptide amino acid sequence, which
CC can be used in the pharmaceutical composition of the invention. The
CC invention relates to a pharmaceutical composition containing at least one
CC activated polyoxalkylene (APO)-modified cationic peptide. The
CC modification of peptides with APO increases their activity against tumour
CC cells, including those with a multidrug resistant phenotype. The
CC pharmaceutical composition can be used to treat tumours, specifically
CC lymphoma, leukaemia, multiple myeloma, or tumours of breast, lung, ovary,
CC cervix, uterus, skin, prostate, liver and colon.

Sequence 42 AA;

alignment_scores:

Quality: 68.00 Length: 58
Ratio: 2.833 Gaps: 2
Percent Similarity: 41.379 Percent Identity: 27.586

alignment_block:

US-09-823-101-2/rev x AAV91698 ..

Align seg 1/1 to: AAV91698 from: 1 to: 42

```

199 GTCAGACACCTCCGATCTGTACACCAACCTTCTCCGCTCCACTCG 150
    :::::|||||
6  IleaArgProPro.....IleaArgProPhetProProhear 20
149 CCCACTTCTACTGCTTGTCTTCTCCAGGTTTGATCTCCCTTCCTC 100
    ||| ::| ||| :::::|
20 gProProIleaArgProProIlehearPro..... 29
99 GGGACTGTAAAGCCATAGTATTGGCTATAGAGCTGACTTATCTCTC 50
    ||| ::| ||| :::::|
30 .....ProIle 31
49 CTACTCTCTTGACACCCCACTC 26
    |||||::: |||||
32 ArgProProhearProProleu 39

```

seq_name: /SID88/gcdata/geneseq/geneseq/AA1996.DAT:AA94447

seq_documentation_block:

ID AAR94447 standard: peptide: 43 AA.

AC AAR94447;

DT 05-NOV-1996 (first entry)

DE Synducin peptide (Bac-5) induces syndecan expression.

KM Synducin; induction; expression; syndecan-1; syndecan-4; surface;
KM mesenchymal cell; fibroblast; epithelial; Bac-5; treatment; stasis;
KM decubitus; ulcers; keloids; skin burns; ischemic tissues;
KM hypercoagulation states; prevention; tumour metastasis; restenosis;
KM inhibition; angiogenesis; proliferation; endothelial.

OS Synthetic.

PN W09609322-A2.

PD 28-MAR-1996.

PF 22-SEP-1995; 95WO-US12080.

PR 22-SEP-1994; 94US-0310722.

PA (CHIL-) CHILDRENS MEDICAL CENT.

PI Bernfield M, Gallo RL;

DR WPI; 1996-188401/19.

PT Modulating mesenchymal interaction by administration of synducin
PT used in the treatment of wounds, tumours, restenosis, etc

PS Claim 4; Pages 26-27; 34pp; English.

CC The present peptide (previously known as an antibacterial peptide)
CC is a synducin, which induces the expression of syndecan-1 and
CC syndecan-4 on the surface of mesenchymal cells, esp. fibroblasts
CC and epithelial cells. Synducins may be used in the treatment of
CC stasis and decubitus ulcers, keloids, skin burns, ischemic tissues
CC and hypercoagulation states, prevention of tumour metastasis,
CC restenosis inhibition and endothelial cell angiogenesis and

CC proliferation induction.
xx
SQ Sequence 43 AA;

alignment_scores:
Quality: 68.00 Length: 58
Ratio: 2.833 Gaps: 2
Percent Similarity: 41.379 Percent Identity: 27.586

alignment_block:
US-09-823-101-2/rev x AAR94447 ..

Align seg 1/1 to: AAR94447 from: 1 to: 43

```
199 GTCAGACACCCCTCCCTGATCTGTGCACCCCAACCTTTCCTGCTGCCCTCCACTCG 150
   ::|||:::|||||   ::   |||||   |||||   ||
  6 IleArgArgProPro.....IleArgProProPheTyrProProPheAr 20

149 CCACCTTCTACTGCGCTTGTCTTCTTCCAGGGTTTGATCTGCCCTTCCTC 100
   ||||   ::   ||   ::|||   ::|||
  20 gProProIleArgProProIlePhePro..... 29

99 GGGACTCTAGCCCATAGTATTGGCTATGAGAGCCTGACTCTTATCTCTC 50
   ::
  30 .....ProIle 31

49 CTACCTCCTTGAGACACCCCACTC 26
   |||||::   |||||
  32 ArgProProPheArgProProIleu 39
```


49 CTACCTCTTGACACCCCACTC 26
 |||||::: |||||
 32 ArgProProPheArgProPheLeu 39

seq_name: /cgn2_6/plodata/2/1aa/PCTUS_COMB.pep:PCT-US95-12080-2

seq_documentation_block:

Sequence 2, Application PC/TUS9512080
 GENERAL INFORMATION:
 APPLICANT: Children's Medical Center Corporation
 TITLE OF INVENTION: Synducin Mediated Modulation of Tissue Repair
 NUMBER OF SEQUENCES: 4
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Patrea L. Pabst
 STREET: 2800 One Atlantic Center
 STREET: 1201 West Peachtree
 CITY: Atlanta
 STATE: Georgia
 COUNTRY: USA
 ZIP: 30309-3450
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: PCT/US95/12080
 FILING DATE:
 CLASSIFICATION:
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (404)-873-8794
 TELEFAX: (404)-815-8795
 INFORMATION FOR SEQ. ID NO.: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 43 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 PCT-US95-12080-2

alignment_scores:
 Quality: 68.00 Length: 58
 Ratio: 2.833 Gaps: 2
 Percent Similarity: 41.379 Percent Identity: 27.586

alignment_block:

US-09-823-101-2/rev x PCT-US95-12080-2 ..

Align seg 1/1 to: PCT-US95-12080-2 from: 1 to: 43

199 GTGAGACACCTCCCTGATCTGTACCCACCTTCTGCTGCCACTCG 150
 :::::|||||
 6 IleArgArgProPro.....IleArgProProPheArgProPheLeu 20
 149 CCACCTCTCTAGCTGTCTCTTCCAGGGTTTGATCTGCCCTTCCTC 100
 |||||::: ||| :::::||||
 20 gProProIleArgProProIlePhePro..... 29
 99 GGGAGCTGTAAAGCATAGTATTGGCTATGAGAGCCTGACTCTTATCTCTC 50
 :::
 30Profile 31

49 CTACCTCTTGACACCCCACTC 26
 |||||::: |||||

32 ArgProProPheArgProPheLeu 39

seq_name: /cgn2_6/plodata/2/1aa/5B_COMB.pep:US-08-343-443B-2

seq_documentation_block:

Sequence 2, Application US/08343443B
 Patent No. 5968734
 GENERAL INFORMATION:

APPLICANT: Aurias, Alain
 APPLICANT: Delattre, Olivier
 APPLICANT: Desmaze, Chantal
 APPLICANT: Melot, Thomas
 APPLICANT: Peter, Marline
 APPLICANT: Ploouastel, Beatrice
 APPLICANT: Thomas, Gilles
 APPLICANT: Zucman, Jessica
 TITLE OF INVENTION: NUCLEIC ACID CORRESPONDING TO A GENE OF
 TITLE OF INVENTION: CHROMOSOME 22 INVOLVED IN RECURRENT CHROMOSOMAL
 TITLE OF INVENTION: TRANSLATIONS ASSOCIATED WITH THE DEVELOPMENT OF CANCEROUS
 TITLE OF INVENTION: TUMORS, AND NUCLEIC ACIDS OF FUSION RESULTING FROM SAID
 NUMBER OF SEQUENCES: 129
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Weiser & Associates
 STREET: 230 South Fifteenth Street
 CITY: Philadelphia
 STATE: PA
 COUNTRY: USA
 ZIP: 19102
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: AEDIT 1.0 DOS text editor
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/343,443B
 FILING DATE: 18-NOV-1994
 CLASSIFICATION: 514
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: PCT/FR93/00494
 FILING DATE: 19-MAY-1993
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: FR 92/06123
 FILING DATE: 20-MAY-1992
 ATTORNEY/AGENT INFORMATION:
 NAME: Weiser, Gerard J.
 REGISTRATION NUMBER: 19,763
 REFERENCE/DOCKET NUMBER: 989,612IP
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 215-875-8383
 TELEFAX: 215-875-8394
 INFORMATION FOR SEQ. ID NO.: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 656 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-343-443B-2

alignment_scores:
 Quality: 66.00 Length: 57
 Ratio: 1.941 Gaps: 2
 Percent Similarity: 59.649 Percent Identity: 36.842

alignment_block:

US-09-823-101-2 x US-08-343-443B-2 ..

Align seg 1/1 to: US-08-343-443B-2 from: 1 to: 656

27 AGTGGGGTGTCCAGAGAGTAGAGAGATAAGTCAGGCTCTCATACG 76
 |||||::: |||||
 282 SerGlyGlyPheSerGlyProGlyGluAsnArgSerMetSer..... 295
 77 CAATACTATAGGCTTACAGTCCCGAGGAAGGAGATCCAAACCTGAA 126
 ||| :::::|||||
 296GlyProAspAsnArgGlyArgGlyArgGlyGlyPheAspA 309
 127 GAAGACAGGCACTAGACTGGCCAGTGGAGGACAGCAAAAGGTTGG... 173
 ||| :::::|||||
 309 rgGlyGlyMetSerArgGlyArgGlyGlyGlyArgGlyGlyMetGly 325

174 ...GTGACAGATCAGGAGG 191
::: |||||
326 SerAlaGlyGluArgGlyGly 332

seq_name: /cgn2_6/ptodata/2/1aa/5A_COMB.pep:US-09-214-564A-4

seq_documentation_block:

; Sequence 4, Application US/09214564A
; Patent No. 6150515
; GENERAL INFORMATION:
; APPLICANT: Sharp, Phillip A.
; APPLICANT: Zhou, Qiang
; TITLE OF INVENTION: TAT-SF: Cofactor For Stimulation Of Transcriptional
; TITLE OF INVENTION: Elongation By HIV-1 TAT
; FILE REFERENCE: M0656/7042
; CURRENT APPLICATION NUMBER: US/09/214,564A
; CURRENT FILING DATE: 1999-08-18
; PRIOR APPLICATION NUMBER: US 60/021,218
; PRIOR FILING DATE: 1996-07-03
; PRIOR APPLICATION NUMBER: US 60/033,152
; PRIOR FILING DATE: 1996-12-13
; PRIOR APPLICATION NUMBER: PCT/US97/11713
; PRIOR FILING DATE: 1997-07-03
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO: 4
; LENGTH: 656
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-214-564A-4

alignment_scores:
Quality: 66.00 Length: 57
Ratio: 1.941 Gaps: 2
Percent Similarity: 59.649 Percent Identity: 36.842

alignment_block:

US-09-823-101-2 x US-09-214-564A-4 ..

Align seg 1/1 to: US-09-214-564A-4 from: 1 to: 656

27 AGTGGGGGTGTCACAGAGGAGTAGAGATTAAGAGTCAGGCTTCATAGC 76
||||||| ::||| ::::::::::|
282 SerGlyGlyPheSerGlyProGlyGluAsnArgSerMetSer..... 295
77 CAATACCTATGCTTACAGTCCCGAGAGAGGAGATCCAAACCTGGAA 126
||| ::::| |||||
296GlyProAspAsnArgGlyArgGlyArgGlyPheAspA 309
127 GAGACAGAGCAGTAGAGTGGCGAGTGGAGCAGCAGAAAGTTGG... 173
|| ||||| ::||| ||||| |||||
309 rGslGlyMetSerArgGlyArgGlyGlyGlyArgGlyGlyMetGly 325
174 ...GTGACAGATCAGGAGG 191
::: |||||
326 SerAlaGlyGluArgGlyGly 332

seq_name: /cgn2_6/ptodata/2/1aa/5A_COMB.pep:US-08-349-696-25

seq_documentation_block:
; Sequence 25, Application US/08349696
; Patent No. 5599671
; GENERAL INFORMATION:
; APPLICANT: Jacobson, Marlene A
; APPLICANT: Johnson, Robert G
; APPLICANT: Luneau, Christopher J
; APPLICANT: Salvatore, Christopher A
; TITLE OF INVENTION: Human Adenosine Receptors
; NUMBER OF SEQUENCES: 28
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Merck & Co., Inc.

STREET: P.O. Box 2000
CITY: Rahway
STATE: NJ
COUNTRY: United States
ZIP: 07065
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: Macintosh IIfx
OPERATING SYSTEM: Macintosh
SOFTWARE: Microsoft Word 5.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/349,696
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: us/08/005945
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Meredith, Roy D.
REGISTRATION NUMBER: 30,777
REFERENCE/DOCKET NUMBER: 186991A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (908)594-4678
TELEFAX: (908)594-4720
INFORMATION FOR SEQ ID NO: 25:
SEQUENCE CHARACTERISTICS:
LENGTH: 318 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-349-696-25

alignment_scores:
Quality: 65.50 Length: 54
Ratio: 2.113 Gaps: 2
Percent Similarity: 57.407 Percent Identity: 31.481

alignment_block:

US-09-823-101-2/rev x US-08-349-696-25 ..

Align seg 1/1 to: US-08-349-696-25 from: 1 to: 318

168 CTTTCGCGTCCCGCCACGCTCTACTGCTGTCTTCACAGG 119
||||| ||||| ::||| ::::::::::|
241 LeuSerTrpLeuProLeuSerIleIleAsnGlyIleIleArgPheAsnG 257
118 TTGGATCTCCCTTCCTCGGAGCTGTAAAGCATATTGGCTATGAG 69
| ::::| |||
257 Y...GluValProGlnLeuValLeu..... 264
68 AGCGTAGCTTATCTCTCTACTCTCTTGAGACACCCACCTACCAAG 19
||| ::||| ||||| ::|||
265TyrMetGlyIleLeuLeuSerHisAlaAsnSerMetMet 277
18 CAGCAGTGGTG 7
::||| ::|||
278 AsnProIleVal 281

seq_name: /cgn2_6/ptodata/2/1aa/5A_COMB.pep:US-08-233-009-25

seq_documentation_block:
; Sequence 25, Application US/08233009
; Patent No. 5646156
; GENERAL INFORMATION:
; APPLICANT: Jacobson, Marlene A
; APPLICANT: Johnson, Robert G
; APPLICANT: Luneau, Christopher J
; APPLICANT: Salvatore, Christopher A
; TITLE OF INVENTION: INHIBITION OF EOSINOPHIL
; TITLE OF INVENTION: ACTIVATION THROUGH A3 ADENOSINE RECEPTOR ANTAGONISM
; NUMBER OF SEQUENCES: 56
; CORRESPONDENCE ADDRESS:

ADDRESSEE: Merck & Co., Inc.
STREET: P.O. Box 2000
CITY: Rahway
STATE: New Jersey
COUNTRY: United States
ZIP: 07065
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/233,009
FILING DATE: 25-Apr-1994
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Bence, Gerard H.
REGISTRATION NUMBER: 35,746
REFERENCE/DOCKET NUMBER: 19219
TELECOMMUNICATION INFORMATION:
TELEPHONE: (908) 594-3901
TELEFAX: (908) 594-4720
INFORMATION FOR SEQ ID NO: 25:
SEQUENCE CHARACTERISTICS:
LENGTH: 318 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
US-08-233-009-25

alignment_scores:
Quality: 65.50 Length: 54
Ratio: 2.113 Gaps: 2
Percent Similarity: 57.407 Percent Identity: 31.481

alignment_block:

US-09-823-101-2/rev x US-08-233-009-25 ..

Align seg 1/1 to: US-08-233-009-25 from: 1 to: 318

```
168 CTTTCCTGCTCCCTCCACCTGCGCCACTTCTACTGCTGCTCTTCCAGG 119
||||| |||||||: :|||:|||||:|||||
241 LeuSerTrpLeuProLeuSerIleIleAsnGlyIleIleTyrrPheAsnG1 257
118 TTGGAGTCGCCCTCTCTCGGAGCTGTAGCCATAGATTGGCTATGAG 69
| :|||:| |
257 Y...GluValProGlnLeuValLeu..... 264
68 AGCCTGACTTATCTCTCTACTCTCTTGAGACACCCCACTCACAAGT 19
|||:|:|:| | |||:|:|:| | :|||
265 .....TyrMetGlyLeuLeuSerHisAlaAsnSerMetMet 277
18 CAGCCAGTGTG 7
::|:|:|:|:|
278 AsnProIleVal 281
```

seq_name: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:US-08-560-231-25

seq_documentation_block:

; Sequence 25, Application US/08560231
; Patent No. 5817760
; GENERAL INFORMATION:
; APPLICANT: Jacobson, Marlene A
; APPLICANT: Johnson, Robert G
; APPLICANT: Luneau, Christopher J
; APPLICANT: Salvatore, Christopher A
; TITLE OF INVENTION: Human Adenosine Receptors
; NUMBER OF SEQUENCES: 28
; CORRESPONDENCE ADDRESS:

ADDRESSEE: Merck & Co., Inc.
STREET: P.O. Box 2000
CITY: Rahway
STATE: NJ
COUNTRY: United States
ZIP: 07065
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: Macintosh IIcx
OPERATING SYSTEM: Macintosh
SOFTWARE: Microsoft Word 5.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/560,231
FILING DATE:
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Meredith, Roy D.
REGISTRATION NUMBER: 30,777
REFERENCE/DOCKET NUMBER: 186991A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (908) 594-4678
TELEFAX: (908) 594-4720
INFORMATION FOR SEQ ID NO: 25:
SEQUENCE CHARACTERISTICS:
LENGTH: 318 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-560-231-25

alignment_scores:
Quality: 65.50 Length: 54
Ratio: 2.113 Gaps: 2
Percent Similarity: 57.407 Percent Identity: 31.481

alignment_block:

US-09-823-101-2/rev x US-08-560-231-25 ..

Align seg 1/1 to: US-08-560-231-25 from: 1 to: 318

```
168 CTTTCCTGCTCCCTCCACCTGCGCCACTTCTACTGCTGCTCTTCCAGG 119
||||| |||||||: :|||:|||||:|||||
241 LeuSerTrpLeuProLeuSerIleIleAsnGlyIleIleTyrrPheAsnG1 257
118 TTGGAGTCGCCCTCTCTCGGAGCTGTAGCCATAGATTGGCTATGAG 69
| :|||:| |
257 Y...GluValProGlnLeuValLeu..... 264
68 AGCCTGACTTATCTCTCTACTCTCTTGAGACACCCCACTCACAAGT 19
|||:|:|:| | |||:|:|:| | :|||
265 .....TyrMetGlyLeuLeuSerHisAlaAsnSerMetMet 277
18 CAGCCAGTGTG 7
::|:|:|:|:|
278 AsnProIleVal 281
```

seq_name: /cgn2_6/ptodata/2/1aa/6B.COMB.pep:US-09-080-704A-25

seq_documentation_block:

; Sequence 25, Application US/09080704A
; Patent No. 6166181
; GENERAL INFORMATION:
; APPLICANT: Jacobson, Marlene A
; APPLICANT: Johnson, Robert G
; APPLICANT: Luneau, Christopher J
; APPLICANT: Salvatore, Christopher A
; TITLE OF INVENTION: Human Adenosine Receptors
; NUMBER OF SEQUENCES: 28
; CORRESPONDENCE ADDRESS:
ADDRESSEE: Merck & Co., Inc.
STREET: P.O. Box 2000

CITY: Rahway
STATE: NJ
COUNTRY: United States
ZIP: 07065
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: Windows NT
SOFTWARE: Word 97
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/080,704A
FILING DATE: 18 May 1998
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Parr, Richard S.
REGISTRATION NUMBER: 32,586
REFERENCE/DOCKET NUMBER: 18699DB
TELECOMMUNICATION INFORMATION:
TELEPHONE: (732)594-4958
TELEFAX: (732)594-4720
INFORMATION FOR SEQ ID NO: 25:
SEQUENCE CHARACTERISTICS:
LENGTH: 318 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-080-704A-25

alignment_scores:
Quality: 65.50 Length: 54
Ratio: 2.113 Gaps: 2
Percent Similarity: 57.407 Percent Identity: 31.481

alignment_block:

US-09-823-101-2/rev x US-09-080-704A-25 ..

Align seg 1/1 to: US-09-080-704A-25 from: 1 to: 318

168 CTTTCCTGCTCCGACCTCCGACCTCTCTGCTCTCTCCAGG 119
||||| |||||||: :|||:|||||:|||||
241 LeuSerTirleuProleuSerIleIleAsnCySyleIleTyrPheAsnG 257
118 TTGGATCTCCCTTCCTCGGACGTAAAGCATAGTATTGGCTATGAG 69
| :|||:|||| |||
257 Y...GluValProGlnIleuValIeu..... 264
68 AGCCTGACTCTTATCTCTCTACCTCTTGAGACACCCCACTCAGCAAGT 19
|||||: ||||||:||||| :|||
265TyrMetGlyIleLeuIleuSerHisAlaAsnSerMetMet 277
18 CAGCAGTGTG 7
278 AsnProIleVal 281

seq_name: /cgn2_6/plodata/2/laa/5B_COMB.pep:US-08-343-443B-107

seq_documentation_block:

; Sequence 107, Application US/08343443B
; Patent No. 5968734
; GENERAL INFORMATION:
; APPLICANT: Aurias, Alain
; APPLICANT: Delattre, Olivier
; APPLICANT: Desmaze, Chantal
; APPLICANT: Melot, Thomas
; APPLICANT: Peter, Martine
; APPLICANT: Ploougaestel, Beatrice
; APPLICANT: Thomas, Gilles
; APPLICANT: Zucman, Jessica
; TITLE OF INVENTION: NUCLEIC ACID CORRESPONDING TO A GENE OF
; CHROMOSOME 22 INVOLVED IN RECURRENT CHROMOSOMAL
; TITLE OF INVENTION: TRANSLATIONS ASSOCIATED WITH THE DEVELOPMENT OF CANCEROUS

TITLE OF INVENTION: TUMORS, AND NUCLEIC ACIDS OF FUSION RESULTING FROM SAID
; TRANSLOCATIONS
; TITLE OF INVENTION: 129
; NUMBER OF SEQUENCES: 129
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Weiser & Associates
; STREET: 230 South Fifteenth Street
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19102
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: AEDIT 1.0 DOS text editor
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/343,443B
FILING DATE: 18-NOV-1994
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/FR93/00494
FILING DATE: 19-MAY-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 92/06123
FILING DATE: 20-MAY-1992
ATTORNEY/AGENT INFORMATION:
NAME: Weiser, Gerard J.
REGISTRATION NUMBER: 19,763
REFERENCE/DOCKET NUMBER: 989.6121P
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-875-8383
TELEFAX: 215-875-8394
INFORMATION FOR SEQ ID NO: 107:
SEQUENCE CHARACTERISTICS:
LENGTH: 295 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-343-443B-107

alignment_scores:
Quality: 65.00 Length: 48
Ratio: 2.241 Gaps: 1
Percent Similarity: 60.417 Percent Identity: 39.583

alignment_block:

US-09-823-101-2 x US-08-343-443B-107 ..

Align seg 1/1 to: US-08-343-443B-107 from: 1 to: 25

27 AGTGGGGGTGTCCAGAGGTAGAGATAGAGATAGAGAGGCTC 76
|||||: |||:|||||:|||||:|||||
46 SerGlyGlyPheSerGlyProGlyIleuAsnArgSerMetSer ... 59
77 CAAATACTATGGCTTACAGTCCGAGAGAGGAGATCCAAACCCCTGGAA 126
||| :|||:|||||:||||| :|||
60GlyProAspAsnArgIleArgIleArgIleArgIlePheAspA 73
127 GAAGACAAAGCACTAGAGAGTGGCCAGTGGAGGCGCAGAAAGGT 170
|| |||||:||||| ||||| |||
73 rgIlyGlyMetSerArgIleArgIleArgIleArgIleArgIleGly 87

seq_name: /cgn2_6/plodata/2/laa/6B_COMB.pep:US-09-540-245A-15

seq_documentation_block:

; Sequence 15, Application US/09540245A
; Patent No. 6270984
; GENERAL INFORMATION:
; APPLICANT: Goodman, Corey
; APPLICANT: Kid, Thomas
; APPLICANT: Brose, Katja
; APPLICANT: Tessier-Lavigne, Marc

```

; TITLE OF INVENTION: Modulating Robo: Ligand Interactions
; FILE REFERENCE: B98-031-3
; CURRENT APPLICATION NUMBER: US/09/540,245A
; CURRENT FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 60/065,544
; PRIOR FILING DATE: 1997-11-14
; PRIOR APPLICATION NUMBER: 60/081,057
; PRIOR FILING DATE: 1998-04-07
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 15
; LENGTH: 1395
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
; US-09-540-245A-15

alignment_scores:
  Quality: 64.00      Length: 64
  Ratio: 1.882      Gaps: 4
  Percent Similarity: 53.125      Percent Identity: 34.375

alignment block:
US-09-823-101-2/rev x US-09-540-245A-15 ..

Align seq 1/1 to: US-09-540-245A-15 from: 1 to: 1395

176 CACCACCTTTCCTCCCTCCACCTG.....CCCACTTACTGCG 136
||||| ||| :::::|||||:|||||:
2 HisPromethisProgluAsnHisAlaIleAlaArgSerThrSerThr 18
135 CTGTGCT.....TCTTCAGGCTTGGATCTCCCT.... 105
: :: |||||:|||||: |||
18 rAsnAsnProSerArgSerArgSerArgMetTrpLeuProLat 35
104 .....TCTCGGAGCTGTAGCCATGATTT 78
||||:||||| |||:
35 rpleuLeuValLeuValAlaSerAsnGlyLeuProAlaValArgLy 51
77 GGCATAGAGCCTGACTTATCTCT...CCTACCTGCTTG 39
||||:||||| :::::|||||:|||||:
52 GlntyrGlnSerProArgIleIleGluHisProThrAspLeu 65

seq_name: /cgn2_6/ptodata/2/1aa/5A_COMB.pep:US-08-362-670B-30

seq_documentation_block:
; Sequence 30, Application US/08362670B
; Patent No. 5658862
; GENERAL INFORMATION:
; APPLICANT: Celeste, Anthony J.
; APPLICANT: Mooney, John
; APPLICANT: Rosen, Vicki A.
; APPLICANT: Wolman, Neil
; APPLICANT: Thomsen, Gerald H.
; APPLICANT: Melton, Douglas A.
; TITLE OF INVENTION: TENDON-INDUCING COMPOSITIONS
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENETICS INSTITUTE, INC.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/362,670B
; FILING DATE: December 22, 1994
; CLASSIFICATION: 514

```

```

; ATTORNEY/AGENT INFORMATION:
; NAME: Lazar, Steven R.
; REGISTRATION NUMBER: 32,618
; REFERENCE/DOCKET NUMBER: 5202-D
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617 498-8260
; TELEFAX: 617 876-5851
; INFORMATION FOR SEQ ID NO: 30:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 240 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-362-670B-30

alignment_scores:
  Quality: 63.50      Length: 63
  Ratio: 1.924      Gaps: 1
  Percent Similarity: 52.381      Percent Identity: 26.984

alignment block:
US-09-823-101-2 x US-08-362-670B-30 ..

Align seq 1/1 to: US-08-362-670B-30 from: 1 to: 240

30 GGGGTGTCAGAGAGTAGAGATAGAGTCAGCTCTCAT..... 74
||||| ||||| :::::|||||:
29 GluGlyAspGlyGlyGlyThrAla***GluGluArgAlaLeuValI 45
75 .....GCCA 78
|||:
45 eserSerArgThrGlnArgGlySerLeuPheArgGluIleArgAlaG 62
:::|||||: |||:
79 AATACATGCTTACAGTCCCGAGAGAGGAGAGATCCAAACCTGGA 128
||||:||||| :::::|||||:
62 IlnAlaArgAlaLeuArgAlaAlaAlaGluProProAspProGly 78
129 AGACAAGGCAGTAGAGTGCGGAGAGGAGGAGGAGGAA 167
|||||:|||||:|||||:|||||:
79 GlyAlaGlySerArgGlyAlaAsnLeuGlyGlyArgArg 91

seq_name: /cgn2_6/ptodata/2/1aa/6A_COMB.pep:US-08-333-576C-30

seq_documentation_block:
; Sequence 30, Application US/08333576C
; Patent No. 6027919
; GENERAL INFORMATION:
; APPLICANT: Celeste, Anthony J.
; APPLICANT: Mooney, John
; APPLICANT: Rosen, Vicki A.
; APPLICANT: Wolman, Neil
; APPLICANT: Thomsen, Gerald H.
; APPLICANT: Melton, Douglas A.
; TITLE OF INVENTION: TENDON-INDUCING COMPOSITIONS
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENETICS INSTITUTE, INC.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/333,576C
; FILING DATE: No. 6027919ember 2, 1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:

```

NAME: Lazar, Steven R.
REGISTRATION NUMBER: 32,618
REFERENCE/DOCKET NUMBER: 5202-B
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617 498-8260
TELEFAX: 617 876-5851
INFORMATION FOR SEQ ID NO: 30:
SEQUENCE CHARACTERISTICS:
LENGTH: 240 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-333-576C-30

alignment_scores:
Quality: 63.50 Length: 63
Ratio: 1.924 Gaps: 1
Percent Similarity: 52.381 Percent Identity: 26.984

alignment_block:

US-09-823-101-2 x US-08-333-576C-30 ..

Align seg 1/1 to: US-08-333-576C-30 from: 1 to: 240

30 GGGGGTGTCCAGAGAGTACAGATAGAGTCTCATATA..... 74
||||| ||||| :|||||
29 GYGLYASPGLYGLYThraAla**GlUGluArGAlaLeuValI 45
75GCCA 78
45 eSerSerArGThrGlnArGlyGluSerLeuPheArGluIleArGAlaG 62
79 AATACTATGCTTACATCCGAGAGGAGATCCAAACCTCGAGA 128
:::|||||:||||| |||:|||||
62 InAlArGAlaLeuArGAlaAlaIaIaGluProProAspProGlyPro 78
129 AGACAAGCAGTAGAAGTGGCGAGTGGAGGAGCAGGAA 167
|||||:|||||:|||||:|||||:|||||
79 GYAlaGlySerArGlyAlaAlaSnLeuGlyGlyArGArg 91

seq_name: /cgn2_6/ptodata/2/1aa/6B_COMB.pep:US-08-808-324-30

seq_documentation_block:

Sequence 30, Application US/08808324
Patent No. 6284872
GENERAL INFORMATION:
APPLICANT: Celeste, Anthony J.
APPLICANT: Wozeney, John
APPLICANT: Rosen, Vicki A.
APPLICANT: Wolfman, Neil
APPLICANT: Thomsen, Gerald H.
APPLICANT: Melton, Douglas A.
TITLE OF INVENTION: TENDON-INDUCING COMPOSITIONS
NUMBER OF SEQUENCES: 35
CORRESPONDENCE ADDRESS:
ADDRESSEE: GENETICS INSTITUTE, INC.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: USA
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/808,324
FILING DATE: Herewith
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Lazar, Steven R.

REGISTRATION NUMBER: 32,618
REFERENCE/DOCKET NUMBER: 5202-D
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617 498-8260
TELEFAX: 617 876-5851
INFORMATION FOR SEQ ID NO: 30:
SEQUENCE CHARACTERISTICS:
LENGTH: 240 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-808-324-30

alignment_scores:
Quality: 63.50 Length: 63
Ratio: 1.924 Gaps: 1
Percent Similarity: 52.381 Percent Identity: 26.984

alignment_block:

US-09-823-101-2 x US-08-808-324-30 ..

Align seg 1/1 to: US-08-808-324-30 from: 1 to: 240

30 GGGGGTGTCCAGAGAGTACAGATAGAGTCTCATATA..... 74
||||| ||||| :|||||
29 GYGLYASPGLYGLYThraAla**GlUGluArGAlaLeuValI 45
75GCCA 78
45 eSerSerArGThrGlnArGlyGluSerLeuPheArGluIleArGAlaG 62
79 AATACTATGCTTACATCCGAGAGGAGATCCAAACCTCGAGA 128
:::|||||:||||| |||:|||||
62 InAlArGAlaLeuArGAlaAlaIaIaGluProProAspProGlyPro 78
129 AGACAAGCAGTAGAAGTGGCGAGTGGAGGAGCAGGAA 167
|||||:|||||:|||||:|||||:|||||
79 GYAlaGlySerArGlyAlaAlaSnLeuGlyGlyArGArg 91

seq_name: /cgn2_6/ptodata/2/1aa/PCTUS_COMB.pep:PCT-US94-14030A-30

seq_documentation_block:

Sequence 30, Application PC/TUS9414030A
GENERAL INFORMATION:
APPLICANT: GENETICS INSTITUTE, INC.
APPLICANT: PRESIDENT AND FELLOWS OF HARVARD COLLEGE
TITLE OF INVENTION: TENDON-INDUCING COMPOSITIONS
NUMBER OF SEQUENCES: 35
CORRESPONDENCE ADDRESS:
ADDRESSEE: GENETICS INSTITUTE, INC.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: USA
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/14030A
FILING DATE: Herewith
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/164,103
FILING DATE: 07-DEC-1993
APPLICATION NUMBER: US 08/217,780
FILING DATE: 25-MAR-1994
APPLICATION NUMBER: US 08/333,576
FILING DATE: 02-NOV-1994
ATTORNEY/AGENT INFORMATION:

NAME: Lazar, Steven R.
REGISTRATION NUMBER: 32,618
REFERENCE/DOCKET NUMBER: 5202D-PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617 498-8260
TELEFAX: 617 876-5851
INFORMATION FOR SEQ ID NO: 30:
SEQUENCE CHARACTERISTICS:
LENGTH: 240 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US94-14030A-30

alignment_scores:
Quality: 63.50 Length: 63
Ratio: 1.924 Gaps: 1
Percent Similarity: 52.381 Percent Identity: 26.984

alignment_block:
US-09-823-101-2 x PCT-US94-14030A-30 ..

Align seg 1/1 to: PCT-US94-14030A-30 from: 1 to: 240

30 GGGGCTGCCAAGGAGTACGAGATACAGCTCCTCATATA..... 74
||||| ||||| :|||||
29 GlyIyspGlyGlyGlyThrAla**GlucIuArgAlaLeuValIle 45
75GCCA 78
45 eSerSerArgThrGlnArgLysGlnSerLeuPheArgGluIleArgAlaG 62
79 AATACATGGCTTACAGTCCCGAGAGGAGATCCAAACCTCGAAGA 128
:|||||: :|||
62 InAlaArgAlaLeuArgAlaAlaIaGluProProAspProGlyPro 78
129 AGACAAGGACAGTACAGTGGCGGAGTGGAGGAGCAAGAA 167
|||||: :|||
79 GlyIaGlySerArgLysAlaAsnLeuGlyGlyArgArg 91

seq_name: /cgn2_6/ptodata/2/1aa/6A_COMB.pep:US-08-897-924A-8

seq_documentation_block:

; Sequence 8, Application US/08897924A
; Patent No. 6028058
; GENERAL INFORMATION:
; APPLICANT: Flokiewicz, Robert Z.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR REGULATING
; TITLE OF INVENTION: NUCLEAR TRAFFICKING OF PROTEINS
; NUMBER OF SEQUENCES: 28
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/897,924A
; FILING DATE: 21-JUL-1997
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 200124.403
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900

TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 67 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-08-897-924A-8

alignment_scores:
Quality: 62.50 Length: 30
Ratio: 3.289 Gaps: 1
Percent Similarity: 63.333 Percent Identity: 53.333

alignment_block:

US-09-823-101-2 x US-08-897-924A-8 ..

Align seg 1/1 to: US-08-897-924A-8 from: 1 to: 67

90 TTACAGTCCGAGGAGGAGATCCAAACCTGGAAGACACAGGCGAG 139
||| :||| ||||| ||||| ||| |||
1 LeuGlyAspArgGlyArgGlyArgAlaLeuProGlyGlyArgLeuGlyI 17
140 TAGAAGT.....GGCGAGTGGAGGCGAGG 164
:|||||: :|||
17 YArgGlyArgGlyArgAlaProGlyArgValGlyGlyArg 30

THIS PAGE BLANK (USPTO)

Percent Similarity: 63.636 Percent Identity: 34.545

alignment_block:

US-09-823-101-2/rev x S72880 ..

Align seg 1/1 to: S72880 from: 1 to: 298

```

185 CTGATCTGTACCCACACCTTCCGCGCCACACGCGCCACTTCTACTGC 136
||||| : : : : : : : : : : : : : : : : : : : : : : :
212 LeuIleuAaYgIntRrTyPrVoAlaRghIsAlaProTyrglyLeuAl 228
||| : : : : : : : : : : : : : : : : : : : : : : :
135 CTGTCTCTCTCCAGGTTTGATCTCCCTCCCGGAGGACTGTAAGCCA 86
| || : : : : : : : : : : : : : : : : : : : : : : :
228 aglySerArghThrleuAlarPvAlSerGIntHrAlaser.....AsnI 243
85 TAGTATTTGGCTATGAGAGCCCTTATCTCTCCACCTCCCTGGAC 36
||||| : : : : : : : : : : : : : : : : : : : : : : :
243 leValIleGlyTyAspIeThrThrglyIleProValGluLysValSer 259
35 ACCCCCACTCACCA 21
||||| |||
260 TyrProThrValGln 264

```

seq_name: pIir2:T15667

seq_documentation_block:

hypothetical protein C27H5.3 - *Caenorhabditis elegans*
 C:Species: *Caenorhabditis elegans*
 C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 24-Nov-1999
 C:Accession: T15667
 R:Pauley, A.
 submitted to the EMBL Data Library, July 1995
 A:Description: The sequence of C. elegans cosmid C27H5.
 A:Reference number: Z18386
 A:Accession: T15667
 A>Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-545 <PAU>
 A:Cross-references: EMBL:U14535; NID:9540265; PID:9540269; PIDN:AAC46657.1; CESP:C27H5.3
 A:Experimental source: strain Bristol N2
 C:Genetics:
 A:Gene: CESP:C27H5.3
 A:Introns: 38/2; 52/1; 87/3; 183/1; 202/2; 262/1; 308/2; 511/3
 C:Superfamily: RNA-binding protein EMS; ribonucleoprotein repeat homology

alignment_scores:

Quality:	75.00	Length:	65
Ratio:	2.419	Gaps:	3
Percent Similarity:	47.692	Percent Identity:	38.462

alignment_block:

US-09-823-101-2 x T15667 ..

Align seg 1/1 to: T15667 from: 1 to: 545

```

27 AGTGGGGGTCCAGAGAGTACAGATAGAGTACAGCTCTCATAGC 76
||||| : : : : : : : : : : : : : : : : : : : : : : :
123 serGlyGlnserGly.....Se 130
77 CAATACTATGGCTTACAGTCCCGAGAGGAGGATCCAAAC..... 119
||||| : : : : : : : : : : : : : : : : : : : : : : :
130 rAspProTyrgly...GlnserArghIleGlyTyrglyTyrglyGly 146
120 .....CTGGAAGAGACAGAGGAGGACTAGAGT 146
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
146 lySerArghIleGlyTyrglyTyrglyTyrglyTyrglyTyrgly 162
147 GGGCGAGTGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 191
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
163 GlyTyrglyTyrglyTyrglyTyrglyTyrglyTyrglyTyrgly 177

```

seq_name: pIir2:JQ1663

seq_documentation_block:

hybrid proline-rich protein - maize

C:Species: *Zea mays* (maize)

C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 24-Sep-1999

C:Accession: JQ1663

R:Jose-Estanyol, M.; Ruiz-Avila, L.; Puigdomenech, P.

Plant Cell 4, 413-423, 1992

A:Title: A maize embryo-specific gene encodes a proline-rich and hydrophobic protein.

A:Reference number: JQ1663; MUID:92361259

A:Accession: JQ1663

A:Molecule type: DNA

A:Residues: 1-301 <COS>

A:Cross-references: EMBL:X60432; NID:g433706; PIDN:CAA42959.1; PID:g433707

A:Experimental source: strain W64A

C:Superfamily: hydroxyproline-rich glycoprotein

alignment_scores:

Quality:	71.50	Length:	61
Ratio:	2.750 <td>Gaps:</td> <td>3 </td>	Gaps:	3
Percent Similarity:	42.623 <td>Percent Identity:</td> <td>29.508</td>	Percent Identity:	29.508

alignment_block:

US-09-823-101-2/rev x JQ1663 ..

Align seg 1/1 to: JQ1663 from: 1 to: 301

```

190 CCCTCCGATCTGCACCAACCTTTCCTCCCTCCGACCTTCCT 141
||||| ||| ||| : : : : : : : : : : : : : : : : : :
161 ProPro.....ThrProTyrglyValProProTyrglyVal 174
140 ACTGCTCTGT.....CTCTCCAGGAGTTGGAT 112
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
174 lProProThrProArProserProProTyrglyValProProTyrglyVal 191
111 CTCCCTCTCTCGGAGCTGTAAGCCATGATTTGGCTATGAGAGCCTGA 62
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
191 roProThrProArProserProPro..... 199
61 CTCTATCTCTCCACTCTCTTGAGACCCGCCA 29
::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
200 .....TyrValProProTyrglyValProPro 207

```

seq_name: pIir2:S04536

seq_documentation_block:

embryonic abundant protein, glycine-rich - maize

C:Species: *Zea mays* (maize)

C:Date: 28-Feb-1990 #sequence_revision 28-Feb-1990 #text_change 23-Jul-1999

C:Accession: S04536

R:Gomez, J.; Sanchez-Martinez, D.; Stiefel, V.; Rigau, J.; Puigdomenech, P.; Pages, M.

Nature 334, 262-264, 1988

A:Title: A gene induced by the plant hormone abscisic acid in response to water stress

A:Reference number: S04536; MUID:88288401

A:Accession: S04536

A:Molecule type: DNA

A:Residues: 1-157 <GOM>

A:Cross-references: EMBL:X12564; NID:g22312; PIDN:CAA31077.1; PID:g22313

C:Genetics:

A:Introns: 38/3

C:Superfamily: glycine-rich RNA-binding protein; ribonucleoprotein repeat homology

F:9-76/Domain: ribonucleoprotein repeat homology <RML>

alignment_scores:

Quality:	70.50	Length:	53
Ratio:	2.431 <td>Gaps:</td> <td>1 </td>	Gaps:	1
Percent Similarity:	54.717 <td>Percent Identity:</td> <td>37.736</td>	Percent Identity:	37.736

alignment_block:

US-09-823-101-2 x S04536 ..


```

235 .....|
109 GAGATCCAAACCTGGAAGAGACAGAGCTGCTGATGACCAATACATATGG 88
242 lyan.....|
159 GGCAGGAAA 167
256 GlyLysLys 258

```

seq_name: p1r2:E96576

seq_documentation_block:

unknown protein, 43598-45751 [imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #extl_change 31-Mar-2001
 C:Accession: E96576
 R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
 Chin, C.W.; Hughes, B.; Huizar, L.
 Nature 408, 816-820, 2000
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
 C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Malt, R.; Marziani,
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
 ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
 A>Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
 A:Reference number: A86141; MUID:21016719
 A:Accession: E96576
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-523 <STO>
 A:Cross-References: GB:AE005173; NID:g10645365; PIDN:AG21485.1; GSPDB:GN00141
 C:Genetics:
 A:Gene: F22610.8
 A:Map position: 1

alignment_scores:
 Quality: 67.50 Length: 50
 Ratio: 2.250 Gaps: 1
 Percent Similarity: 60.000 Percent Identity: 36.000

alignment_block:
 US-09-823-101-2 x E96576 ..

Align seg 1/1 to: E96576 from: 1 to: 523

```

39 CAAGAGGTAGAGAGATAAGAGTCAAGCTCTCATGACCAATACATATGG 88
272 G1uAlaG1yArGArGAlaArGSeRgluLeuSeRArglYgluAlaGluG1 288
89 CTTCACGTCCCGAGAGGAGGAGATCCAAACCTGGAAGAGACAGGCA 138
288 ySeRSeRValglYglYArGglYArGglYArGglYArGglYArGglY 305
139 GTAGAGT...GGCAGAGTGGAGCAGAGAAAGTTGGTGACAGATCAG 185
305 lArGglYArGglYArGglYArGglYArGglYArGglYArGglYArGglY 321

```

seq_name: p1r2:A36502

seq_documentation_block:

insulin receptor-related receptor precursor - guinea pig
 C:Species: Cavia porcellus (guinea pig)
 C:Date: 01-Feb-1991 #sequence_revision 01-Feb-1991 #extl_change 23-May-1997
 C:Accession: A36502
 R:Shier, P.; Malt, V.M.
 J. Biol. Chem. 264, 14605-14608, 1989
 A>Title: Primary structure of a putative receptor for a ligand of the insulin family.
 A:Reference number: A36502; MUID:89359245
 A:Accession: A36502

A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1300 <SH1>
 A:Cross-References: GB:J05047
 C:Superfamily: Insulin receptor; protein kinase homology
 C:Keywords: ATP; autophosphorylation; phosphoprotein; receptor; transmembrane protein
 F:977-1253/Domain: protein kinase homology <KIN>
 F:985-993/Region: protein kinase ATP-binding motif

alignment_scores:
 Quality: 67.50 Length: 79
 Ratio: 1.731 Gaps: 4
 Percent Similarity: 49.367 Percent Identity: 29.114

alignment_block:

US-09-823-101-2/rev x A36502 ..

Align seg 1/1 to: A36502 from: 1 to: 1300

```

195 GACACCCCTCCCTGATCTGTACCCACCTTCCGCTCCACCTCGGCCA 146
92 GluSerLeuArGAspLeuPheProAsnLeuAlaValIleArgG1yAla1 108
145 CTTC.....TACTGCTTGTCTTCCAG.....G 120
108 sleuPheLeuGlYtYrAlaLeuValIlePheGluMetProHisLeuArG 125
119 GTTGGATCTCCCTCTCCGAGCTGTAAGCCATAGTAT..... 79
125 sPValglYleuProAlaLeuGlYAlaValleuHisG1SerValArGVal 141
78 .....TGG...CTATG 71
142 GluLysAsnGlnGluLeuCysHisLeuSerThrIleAspTrpGlyLeu 158
70 AGAGCTGACTTATCTCTCTACCTCTTGGACAC 34
158 uGlnProThrProSerThrAsnTrpIleValGlyAsn 170

```


THIS PAGE BLANK (USPTO)

OM of: US-09-823-101-2 to: SwissProt_39.* out_format : pfs

Date: Jan 17, 2002 4:08 PM

About: Results were produced by the Gencore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

```
-MODE=frame+n2p.model -DEV=xlp
-Q/cgn2.1/USPTO.spool/US09823101/runat.17012002.155138.22061/app-query.fasta.1.2239
-DB=SwissProt_39 -OEM=fastan -SUFFIX=n2p.rsp -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPEXT=0.000 -LOOPEXT=0.000
-GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FCAPOP=6.000 -FCAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blomsu62
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR.SCORE=pct
-THR.MAX=100 -THR.MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFM=pfs
-NORMEXT=-MINLEN=0 -MAXLEN=2000000000
-USER=US09823101@cgn1.1.0 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
-MAIT -THREADS=1
```

Search information block:

```
Query: US-09-823-101-2
Query length: 235
Database: SwissProt_39.*
Database sequences: 100059
Database length: 36664827
Search time (sec): 66.470000
```

score_list:

Sequence	Strd Orig	Zscore	EScore	len	Documentation
SwissProt_39:GRPA_MAIZE	70.50	136.78	3.85	157	P10979 zea mays (maize). g1y
SwissProt_39:YASX_DROME	70.50	124.91	4.19	661	P09052 drosophila melanogast
SwissProt_39:BCIT_BOVIN	68.00	130.79	7.39	176	P19660 bos taurus (bovine)
SwissProt_39:IRRC_CAVO	67.50	113.27	9.47	1300	P14617 cavia porcellus (gult)
SwissProt_39:SMDS_DROME	67.00	130.04	10.63	151	O4437 drosophila melanogast
SwissProt_39:SUMA_DROME	67.00	114.09	9.49	1042	P12297 drosophila melanogast
SwissProt_39:SCBP_SCHPO	66.50	123.23	11.26	305	P35551 schizosaccharomyces f
SwissProt_39:ARPI_HSVIF	66.50	122.33	11.33	340	P17588 herpes simplex virus
SwissProt_39:ROG_MOUSE	66.50	121.24	11.42	388	O35479 mus musculus (mouse)
SwissProt_39:ROG_HUMAN	66.50	121.18	11.42	391	P38159 homo sapiens (human)
SwissProt_39:RS2_ARATH	66.00	122.78	12.76	285	P49688 arabidopsis thaliana
SwissProt_39:RSG_HUMAN	66.00	115.90	13.40	656	O10844 homo sapiens (human)
SwissProt_39:RA3R_HUMAN	65.50	120.87	14.62	318	P37365 homo sapiens (human)
SwissProt_39:RBR1_YEAST	65.50	120.64	14.64	327	P15646 saccharomyces cerevis
SwissProt_39:FUS_BOVIN	65.50	116.93	15.03	512	O28009 bos taurus (bovine)
SwissProt_39:HEV1_YARLI	65.50	114.70	15.27	671	P87200 yarrowia lipolytica
SwissProt_39:GAG_FSVST	64.50	117.58	19.10	371	P03338 feline sarcoma virus
SwissProt_39:GAG_FSVGA	64.50	116.45	19.26	425	P03337 feline sarcoma virus
SwissProt_39:CC12_SCHPO	64.50	116.35	20.99	1841	O10059 schizosaccharomyces
SwissProt_39:FBRL_XENLA	64.00	117.71	21.56	323	P22232 xenopus laevis (afric
SwissProt_39:CIRP_MOUSE	63.50	121.91	23.65	172	O61413 mus musculus (mouse)
SwissProt_39:GAR1_SCHPO	63.50	120.91	23.81	194	O06975 schizosaccharomyces f
SwissProt_39:GAG_FSVH2	63.50	114.65	24.90	414	P04322 feline sarcoma virus
SwissProt_39:RECQ_AOUAE	63.50	112.77	25.24	520	O06834 aquifex aeolicus. dna
SwissProt_39:GAG_FIV	63.50	111.87	25.40	580	P10262 feline leukemia virus. g
SwissProt_39:PAK1_SCHPO	63.50	110.87	25.58	655	O61545 mus musculus (mouse)
SwissProt_39:PAK1_SCHPO	63.50	110.83	25.59	658	P50527 schizosaccharomyces f
SwissProt_39:PAK1_SCHPO	63.50	108.65	25.99	857	O18381 drosophila melanogast
SwissProt_39:RAD1_ARATH	63.50	107.84	26.14	945	P42762 arabidopsis thaliana
SwissProt_39:AGOI_ARATH	63.50	106.99	26.30	1048	O04379 arabidopsis thaliana
SwissProt_39:CEK1_SCHPO	63.50	104.97	26.68	1338	P38938 schizosaccharomyces
SwissProt_39:U2AG_DROME	63.00	117.36	27.60	264	O94535 drosophila melanogast
SwissProt_39:TEB6_HSVIE	63.00	116.50	27.77	293	P28940 equine herpesvirus 1y
SwissProt_39:HEB6_PSEAE	63.00	114.62	28.14	368	O87125 pseudomonas aeruginos
SwissProt_39:EXLP_TOBAC	63.00	113.41	28.39	426	O03211 nicotiana tabacum (cc
SwissProt_39:NUCL_XENLA	63.00	111.67	28.74	526	P35637 homo sapiens (human)
SwissProt_39:SRAL_RAR	63.00	105.98	29.10	650	P20397 xenopus laevis (afric
SwissProt_39:LSMA_CABEL	62.50	122.66	30.03	1048	O63527 rattus norvegicus (rat
				123	O19552 caenorhabditis elegans

```
SwissProt_39:PO61_MOUSE - 62.50 115.27 31.65 301 O07916 mus musculus (mous
SwissProt_39:HEV1_HUMAN - 62.50 115.19 31.67 304 O97513 homo sapiens (huma
SwissProt_39:SAHH_ARATH - 62.50 111.33 32.55 485 O23255 arabidopsis thalia
SwissProt_39:SAHH_MEDSA - 62.50 111.33 32.55 485 P50246 medicago sativa (a
SwissProt_39:SAHH_WHEAT - 62.50 111.33 32.55 485 P32112 triticum aestivum
```

seq_name: SwissProt_39:GRPA_MAIZE

```
seq_documentation_block:
ID GRPA_MAIZE STANDARD: PRT: 157 AA.
```

```
AC P10979;
DT 01-JUN-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE GLYCINE-RICH RNA-BINDING, ABSICISIC ACID-INDUCIBLE PROTEIN.
GN RAB15.
OS Zea mays (Maize).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
OC Panicoideae; Andropogoneae; Zea.
OX NCBI_TaxID=4577;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Endosperm;
RX MEDLINE=88286401; PubMed=2969461;
RA Gomez J., Sanchez-Martinez D., Stiefel V., Rigau J., Puigdomenech P.,
RA Pages M.;
RT "A gene induced by the plant hormone abscisic acid in response to
RT water stress encodes a glycine-rich protein.";
RL Nature 334:262-264(1988).
RN [2]
RP SIMILARITY TO RNA-BINDING PROTEINS.
RX MEDLINE=89097302; PubMed=2521378;
RA Mortenson E., Dreyfuss G.;
RA "RNP in maize protein.";
RL Nature 337:312-312(1989).
```

```
-1- FUNCTION: POSSIBLY HAS A ROLE IN RNA TRANSCRIPTION OR
PROCESSING DURING STRESS.
-1- INDUCTION: BY THE PLANT HORMONE ABSICISIC ACID IN RESPONSE TO
WATER STRESS.
-1- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RRM).
-----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```
EMBL: X12564; CAA31077.1; -.
PIR: S04536; S04536.
HSP: P09651; JUP1.
MatzEB: 69261; -.
InterPro: IPR000504; RRM.
DR Pfam: PF00076; rrm; 1.
DR SMART: SM00360; RRM; 1.
DR PROSITE: PS0102; RRM; 1.
DR PROSITE: PS00030; RRM_RNP_1; 1.
KW RNA-binding.
FT DOMAIN 8 86 RNA-BINDING (RRM).
FT DOMAIN 88 154 GLY-RICH.
SQ SEQUENCE 157 AA; 15438 MW; 8A10592248D60D16 CRC64;
```

alignment_scores:

```
Quality: 70.50 Length: 53
Ratio: 2.431 Gaps: 1
Percent Similarity: 54.717 Percent Identity: 37.736
```

```
alignment_block:
US-09-823-101-2 x GRPA_MAIZE ..
```



```

RP  [3] STRUCTURE BY NMR OF 131-173.
RX  MEDLINE=96183292; PubMed=8605180;
RA  Raj P.A., Marcus E., Edgerton M.;
RT  "Delineation of an active fragment and poly(L-proline) II
RL  conformation for candidal activity of bacitracin 5.";
RN  Biochemistry 35:4314-4325(1996).
RN  [4]
RP  CHARACTERIZATION.
RX  MEDLINE=96300243; PubMed=8706679;
RA  Storici P., Tossi A., Lenarcic B., Romeo D.;
RT  "Purification and structural characterization of bovine
RL  cathelicidins, precursors of antimicrobial peptides.";
RN  Eur. J. Biochem. 238:769-776(1996).
CC  -I- FUNCTION: EXERTS, IN VITRO, A POTENT ANTIMICROBIAL ACTIVITY.
CC  PROBABLY DUE TO AN IMPAIRMENT OF THE FUNCTION OF THE RESPIRATORY
CC  CHAIN AND OF ENERGY-DEPENDENT ACTIVITIES IN THE INNER MEMBRANE
CC  OF SUSCEPTIBLE MICROORGANISMS.
CC  -I- TISSUE SPECIFICITY: LARGE GRANULES OF NEUTROPHILS.
CC  -I- DOMAIN: BACS SEQUENCE CONSISTS ALMOST EXCLUSIVELY OF X-P-P-Y
CC  REPEATS.
CC  -I- PTM: ELASTASE IS RESPONSIBLE FOR ITS MATURATION.
CC  -I- MASS SPECTROMETRY: MW=16570; MW_ERR=1; METHOD=ELECTROSPRAY;
CC  RANGE=30-173.
CC  -I- SIMILARITY: BELONGS TO THE CATHELICIDIN FAMILY.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL: L02650; AAA30404.1; .
DR  PIR: B36589; B36589.
DR  PIR: A45328; A45328.
DR  InterPro: IPR001894; Cathelicidin.
DR  InterPro: IPR000010; Cystatin.
DR  Pfam: PF00666; Cathelicidins; 1.
DR  Prodom: PD001838; Cathelicidin; 1.
DR  SMART: SM00043; Cy; 1.
DR  PROSITE: PS00946; CATHELICIDINS_1; 1.
DR  PROSITE: PS00947; CATHELICIDINS_2; 1.
DR  Antibiotic; Repeat; Signal; Amidation.
FT  SIGNAL 1 29 POTENTIAL.
FT  PROPEP 30 130
FT  PEPTIDE 131 173
FT  PROPEP 174 176 REMOVED IN MATURE FORM.
FT  MOD_RFS 30 30 PYRROLIDONE CARBOXYLIC ACID.
FT  DISULFID 85 96 BY SIMILARITY.
FT  DISULFID 107 124 BY SIMILARITY.
FT  MOD_RES 173 173 AMIDATION (G-174 PROVIDE AMIDE GROUP).
FT  CONFLICT 170 171 GP -> R (IN REF. 2).
SQ  SEQUENCE 176 AA; 20030 MW; DZDI81C059294I5 GPC64;

alignment_scores:
    quality: 68.00      length: 58
    ratio: 2.833        gaps: 2
    percent similarity: 41.379      percent identity: 27.586

alignment_block:
US-09-823-101-2/rev x BCTS_BOVIN ..

Align seg 1/1 to: BCTS_BOVIN from: 1 to: 176

199  GTCACACACCCCTGCTGATCTGTACACCAACCTTCTGCTCCGACATCG 150
    :::::::::::::::  ::  ||||||| ||||||  ||
136  ILEARGARGPROFRO.....ILEARGPROPHETRYFPRIOPROPHAR 150
149  CCCACTTACTGCTGTCTTCTTCCAGGGTTTGAGATCTCCCTTCCTC 100
||||  ::  |||  ::::::::::|||

```


[illegible]

```

DR InterPro: IPR001163; snRNP_Sm.
DR Pfam: PF01423; Sm; 1.
KW Nuclear Protein; Ribonucleoprotein; mRNA splicing;
KW mRNA processing; Repeat.
FT DOMAIN 84 120 ARG/LYS-RICH (BASIC).
FT DOMAIN 108 113 3 X 2 AA TANDEM REPEATS OF RG.
SQ SEQUENCE 151 AA; 15582 MW; 65F0B899EACAlc77 CRC64;

alignment_scores:
    Quality: 67.00      Length: 65
    Ratio: 2.030        Gaps: 2
    Percent Similarity: 50.769      Percent Identity: 33.846

alignment_block:
US-09-823-101-2 x SMD3_DROME ..

Align seg 1/1 to: SMD3_DROME from: 1 to: 151

6 TCACCACTGGCTGACTGGTGAGTGGGGGCTGCCAAGAGGATGAGACAGA 55
   :::::      :::::      :::::
80 AAlapromPhelylsysglInthrglysglyLeuylgYlThraAlaGl 96
   :::::      :::::      :::::
56 TAAGAGTCAGGCTCTCATAGCCAAATFACTAGCCTTAAGATGCCGAGAA 105
   :::::      :::::      :::::
96 yArgglYlsAlaAlaAlaIleLeuArg.....AlaGlAlaAlaAggGlyA 110
   :::::      :::::      :::::
106 GG.....GGAGATCCCAACCTCGAGAGAGACAGACGATGAGAGT 146
   :::::      :::::      :::::
110 rGglYArgglYglYProProglYglYArgglYThrlYglYlYProPro 126
   :::::      :::::      :::::
147 GGGCGAGTGGGAGGAGGAGGAAGTTGGGTGACAGATCAGAGGAGGG 191
   :::::      :::::      :::::
127 GYAlaProglYglYserglYglYArgglYAlaIArgglYnglYglY 141

seq_name: SwissProt_39:SUWA_DROME

seq_documentation_block:
ID SUWA_DROME STANDARD; PRT; 1042 AA.
AC
PI 12297; O24543;
DT 01-OCT-1989 (Rel. 12, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE SUPPRESSOR OF WHITE APRICOT PROTEIN.
GN SU(WA).
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN 11
RP SEQUENCE FROM N.A.
RX MEDLINE=8816654; PubMed=2832151;
RA Chou T.-B., Zachar Z., Bingham P.M.;
RT "Developmental expression of a regulatory gene is programmed at the
RT level of splicing.";
RL EMBO J. 6:4095-4104(1987).
RN 12
RP FUNCTION, AND ALTERNATIVE SPLICING.
RX MEDLINE=8616655; PubMed=3443103;
RA Zachar Z., Chou T.-B., Bingham P.M.;
RT "Evidence that a regulatory gene autoregulates splicing of its
RT transcript.";
RL EMBO J. 6:4105-4111(1987).
RN 13
RP CHARACTERIZATION OF RS DOMAIN.
RX MEDLINE=92005712; PubMed=1655279;
RA Li H., Bingham P.M.;
RT "Arginine/serine-rich domains of the su(wa) and tra RNA processing
RT regulators target proteins to a subnuclear compartment implicated in
RT splicing.";
RL Cell 67:335-342(1991).
CC -1-FUNCTION: REGULATOR OF PRE-MRNA SPLICING (AND, POSSIBLY, OF OTHER

```

```

CC RNAPROCESSING EVENTS). REGULATE ITS OWN EXPRESSION AT THE LEVEL
CC OF RNA PROCESSING.
CC
CC -1- SUBCELLULAR LOCATION: NUCLEAR; SPECKLED SUBNUCLEAR COMPARTMENT.
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
CC SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- DEVELOPMENTAL STAGE: THREE MRNAs ARE PRODUCED DURING DEVELOPMENT.
CC THE SMALLEST OF THESE (3.5 KB RNA) IS THE MAJORITY SPECIES DURING
CC PRECELLULAR BLASTODERM DEVELOPMENT AFTER WHICH ITS LEVELS DROP
CC RAPIDLY, BUT PERSISTS AS A MINORITY SPECIES THROUGHOUT THE REST OF
CC THE LIFE OF THE ORGANISM. THE LARGER TWO TRANSCRIPTS (4.4 AND 5.2
CC KB RNAs) FIRST APPEAR AROUND CELLULAR BLASTODERM AND LEVELS
CC INCREASE SUBSTANTIALLY DURING NEXT FEW HOURS AND ARE THE
CC PREPREDOMINANT RNA SPECIES THROUGHOUT THE REMAINDER OF THE LIFE OF
CC THE ORGANISM.
CC
CC -1- DOMAIN: KS DOMAIN DIRECTS LOCALIZATION OF PROTEINS TO THE SPECKLED
CC SUBNUCLEAR COMPARTMENT AND THE PURPOSE OF THIS LOCALIZATION IS TO
CC ALLOW COLOCALIZATION AND CO-CONCENTRATION OF COMPONENTS OF THE
CC SPLICING AND SPLICING REGULATORY MACHINERY TO PERMIT RELATIVELY
CC HIGH RATES AND/OR EFFICIENCIES OF REACTION AND INTERACTION.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-ib.ch).
CC -----
CC EMBL; X06589; CAA29812.1; -.
CC EMBL; X06589; CAA29813.1; -.
CC PIR; S06028; S06028.
CC FlyBase; FBgn0003638; su(w[al]).
CC InterPro; IPR000061; Surp.
CC Pfam; PF01805; Surp. 2.
CC Transcription regulation; RNA-binding; mRNA splicing; Repeat;
CC KW Nuclear protein; Alternative splicing.
CC
CC FT DOMAIN 47 177 POLY-SER.
CC FT DOMAIN 51 177 DRY CEEERU MOTIF.
CC FT DOMAIN 313 602 2 X REPEATS OF THE SURP MOTIF.
CC FT REPEAT 313 355 SURP MOTIF 1.
CC FT REPEAT 562 602 SURP MOTIF 2.
CC FT DOMAIN 528 533 POLY-GLU.
CC FT DOMAIN 640 643 POLY-ALA.
CC FT DOMAIN 826 829 POLY-ASP.
CC FT DOMAIN 829 964 ARG/SER-RICH (RS DOMAIN).
CC FT DOMAIN 1035 1038 POLY-ARG.
CC FT VARSPLIC 100 177 MISSING (IN SHORT ISOFORM).
CC FT SEQUENCE 1042 AA; 114928 MW; E564006EE163C3B3 CRC64;

```

```

434 GlnAlaValLeuGlnGluAspGluSerSerAspProGlyAsnSerglnH1 450
135 .....GGCAGTAGAAGTGGCGAGTGGAGGACGAGGAAGCTTGGGTGA 177
      |||::: ||| ::::: |||:::
450 sSergIglyThrAlaSerProAlaLeuSerCyArSgSergGluGlyHis 467
178 CAGATCAGGAGGAGGCTGCTGACCTTTTCTTGAGGAATTCATTAGGC 224
      ::::: ||| ||| ::::: |||
467 sSergIglyGlyGluPheThrProValLeuGlnGlyIrrAsnGly 482

seq_name: SwissProt_39:FBRL_SCHPO

seq_documentation_block:
ID   FBRL_SCHPO          STANDARD;          PRT;          305 AA.
AC   P35551;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   FIBRILLARIN.
GN   FIB OR SPBC2D10.10C.
OS   Schizosaccharomyces pombe (fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomycetes.
OX   NCBI_TaxID=4896;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972;
RX   MEDLINE=93261822; PubMed=8493104;
RA   Glard J.-P., Fellu J., Calzeagues-Ferrier M., Lapeyre B.;
RT   "Study of multiple fibrillar mRNAs reveals that 3' end formation in
RN   Schizosaccharomycetes is sensitive to cold shock.";
RN   Nucleic Acids Res. 21:1881-1887(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972;
RA   Wood V., Rajandream M.A., Barrell B.G., Taylor K., Harris D.;
RL   Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
CC   -1- FUNCTION: FIBRILLARIN IS A COMPONENT OF A NUCLEOLAR SMALL NUCLEAR
CC   RIBONUCLEOPROTEIN PARTICLE THOUGHT TO PARTICIPATE IN THE FIRST
CC   STEP IN PROCESSING PRERIBOSOMAL RNA.
CC   -1- SUBCELLULAR LOCATION: NUCLEAR; FIBRILLAR REGION OF THE NUCLEOLUS.
CC   -1- PTM: BY HOMOLOGY TO OTHER FIBRILLARINS, SOME OR ALL OF THE
CC   N-TERMINAL ARGININES ARE N,N-DIMETHYLATED (DMN).
CC   -1- SIMILARITY: BELONGS TO THE FIBRILLARIN FAMILY.
CC   -----
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----
CC   EMBL; X69930; CAA49550.1; -.
CC   EMBL; AL031788; CAA21168.1; -.
CC   PIR; S33690; S33690.
CC   InterPro: IPR000692; FibrillarIn.
DR   Pfam; PF01269; FibrillarIn.1.
DR   PRINTS; PRK0052; FibrillarIn.1.
DR   ProDom; PD004637; FibrillarIn.1.
DR   PROSITE; PS00566; FIBRILLARIN.1.
KW   Nuclear protein; Methylation; Ribonucleoprotein; rRNA processing;
KW   RNA-binding.
FT   DOMAIN            6          70          GLY/DNA-RICH.
SQ   SEQUENCE 305 AA; 32040 MW; 5AE6B8B37EC3331 CRC64;

Alignment_scores:
Quality: 66.50          Length: 54
Ratio: 2.463           Gaps: 2
Percent Similarity: 50.000 Percent Identity: 38.889

```


alignment_block:

US-09-823-101-2 x FBRL_SCHPO ..

Align seg 1/1 to: FBRL_SCHPO from: 1 to: 305

```
30 GGGGGTGTCCAGAGAGATAGAGATAGAGTACAGGCTCTCATAGCCAA 79
||||| :|||
19 GtlyGlyPheAnsnglyGlyArg..... 25
80 ATACTATGGCTTACAGTCCCGAGAGAGGAGATCCAAACCTGGAGAGA 129
||||| :|||
26 .....GtlyGlyPheGlyGlyArgGlyAlaArgGlyGlyGly 40
130 GACAAGCGCAGTAGAGTAGGCGAGCGAGCGCAAGAGTGGTGACA 179
|| :|||
40 rGtlyGlyAlaArgGlyGlyArg...GtlyGlyArgGlyAlaArgGly 55
180 GATCAGCGAGG 191
:::|||||
56 GtlyArgGlyGly 59
```

seq_name: Swissprot_39:LRPL_HSVLF

seq_documentation_block:

```
ID LRPL_HSVLF STANDARD; PRT; 340 AA.
AC P17588;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 01-FEB-1994 (Rel. 28, Last annotation update)
DE LATENCY-RELATED PROTEIN 1.
OS Herpes simplex virus (type 1 / strain F).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Simplexvirus.
OX NCBI_TaxID=10304;
```

```
RM SEQUENCE FROM N.A.
RX MEDLINE=89085598; PubMed=2535901;
RA Wechsler S.L., Nesburn A.B., Zwaagstra J., Ghiasi H.;
RT "Sequence of the latency-related gene of herpes simplex virus type
1."
RL Virology 168:168-172(1989).
```

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

```
DR EMBL: J04323; AAA45799.1; -.
KW PIR: A33337; WMBEL1.
FT REPEAT.
FT DOMAIN 27 75 2 x 17 AA REPEATS.
FT REPEAT 27 43 1.
FT REPEAT 59 75 2.
SQ SEQUENCE 340 AA; 35604 MM; 2977FA8F87E5451 CRC64;
```

alignment_scores:

Quality: 66.50 Length: 44
Ratio: 3.167 Gaps: 1
Percent Similarity: 47.727 Percent Identity: 40.909

alignment_block:

US-09-823-101-2 x LRPL_HSVLF ..

Align seg 1/1 to: LRPL_HSVLF from: 1 to: 340

```
30 GGGGGTGTCCAGAGAGATAGAGATAGAGTACAGGCTCTCATAGCCAA 79
||||| :|||
299 GtlyGlyAlaArgGlyGlyProArgGly..... 307
```

```
80 ATACTATGGCTTACAGTCCCGAGAGAGGAGATCCAAACCTGGAGAGA 129
||||| :|||
308 .....SerArgGlyArgGlyGlyArgGlyArgGlyGly 319
130 GACAAGCGCAGTAGAGTAGGCGGAGCGAGCGAGG 161
|| :|||
319 rGtlyGlyAlaArgArgGlyArgGlyGlyGly 329
```

seq_name: Swissprot_39:ROG_MOUSE

seq_documentation_block:

```
ID ROG_MOUSE STANDARD; PRT; 388 AA.
AC O35479;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HETEROGENEOUS NUCLEAR RIBONUCLEOPROTEIN G (HNRP G).
GN RBMXPI OR RBMXRT OR HNRPG OR HNRNPg.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
```

```
RM SEQUENCE FROM N.A.
RX MEDLINE=98119027; PubMed=9457683;
RA Delbridge M.L., Ma K., Subbarao M.N., Cooke H.J., Bhasin S.,
RA Graves J.A.M.;
RT "Evolution of mammalian HNRPg and its relationship with the putative
azooaspermatid factor RBM."
RL Mamm. Genome 9:168-170(1998).
CC -1- FUNCTION: BINDS RNA.
CC -1- SUBCELLULAR LOCATION: NUCLEAR; COMPONENT OF RIBONUCLEOSOMES.
CC -1- PTM: O-GLYCOSYLATED (BY SIMILARITY).
CC -1- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RRM).
```

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

```
DR EMBL: AF031568; AAB86639.1; -.
DR MGD: MG1:1343045; Rbmxt.
DR InterPro: IPR000504; RRM.
DR Pfam: PF00076; Rrm; 1.
DR SMART: SM00360; RRM; 1.
DR PROSITE: PS00102; RRM; 1.
DR PROSITE: PS00030; RRM_RNP_1; 1.
FT Nucleic acid binding; RNA-binding; Ribonucleoprotein; Glycoprotein.
FT DOMAIN 8 86
SQ SEQUENCE 388 AA; 42233 MM; F02F2DCB8C87F14 CRC64;
```

alignment_scores:

Quality: 66.50 Length: 68
Ratio: 2.375 Gaps: 2
Percent Similarity: 41.176 Percent Identity: 32.353

alignment_block:

US-09-823-101-2 x ROG_MOUSE ..

Align seg 1/1 to: ROG_MOUSE from: 1 to: 388

```
42 GGAGGTAGGAGAGATAGAGTACAGGCTCTCATAGCCAAATACTAT..... 86
||||| :|||
318 GtlyGlySerArgAspSerArgSerArgSerArgSerArgSerArgSerArg 334
87 .....GGCTTACAGTCCCGAGGA..... 104
334 rGtlyArgAspArgValGtlyArgGlnGlnArgGlyLeuProProSerMetG 351
```

105AGGGA 110
 351 LuArgGLYTYRProProRoArgAspSerTYRserSerSerArgGLY 367
 111 GATCAAAACCTCGAGAGACAGACGAGTGAAGTGGCGAGTGGAGG 160
 368 AlaProArgGLYGLYGLYArgGLYGLYSerArgSerArgGLYGLY 384
 161 CAGG 164
 384 YArg 385

seq_name: SwissProt_39:ROG_HUMAN

seq_documentation_block:
 ID ROG_HUMAN STANDARD; PRT; 391 AA.
 AC P38159;
 DT 01-OCT-1994 (Rel. 30, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE HETEROGENEOUS NUCLEAR RIBONUCLEOPROTEIN G (HNRP G) (GLYCOPROTEIN P43).
 GN RBMX1 OR HNRP.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Breast;
 MEDLINE=94021365; PubMed=7692398;
 RA Soulaud M., della Valle V., Stomi M., Pinol-Roma S., Codogno P., Bauvy C., Belli M., Lacroix J.-C., Monod G., Dreyfuss G., Larsen C.-J.,
 RT "hnRNP G: sequence and characterization of a glycosylated RNA-binding protein.";
 RT Nucleic Acids Res. 21:4210-4217(1993).
 RL (2)
 RA REVISIONS.
 RA Venables J.P., Larsen C.-J.;
 RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.

CC
 CC -1- FUNCTION: BINDS RNA.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR; COMPONENT OF RIBONUCLEOSOMES.
 CC -1- PTM: O-GLYCOSYLATED.
 CC -1- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RRM).
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC
 CC EMBL: Z23064; CAA80599.1;
 CC DR PIR: S43202; S43202.
 CC DR HSP: P09651; IUP1.
 CC DR MIM: 300199;
 CC DR InterPro: IPR000504; RRM.
 CC DR Pfam: PF00076; RRM_1.
 CC DR SMART: SM00360; RRM; 1.
 CC DR PROSITE: PS50102; RRM; 1.
 CC DR PROSITE: PS00030; RRM_RNP_1; 1.
 CC KW Nuclear protein; RNA-binding; Ribonucleoprotein; Glycoprotein.
 CC FT DOMAIN 8
 CC SEQUENCE 391 AA; 42403 MW; 904FE911FBD3626 CRC64;

alignment_scores:
 Quality: 66.50 Length: 68
 Ratio: 2.375 Gaps: 2
 Percent Similarity: 41.176 Percent Identity: 32.353

alignement_block:
 US-09-823-101-2 x ROG_HUMAN ..
 Align seg 1/1 to: ROG_HUMAN from: 1 to: 391

42 GGAGTAGAGAGATAAGAGTACAGCTCATAGCCAAATACAT..... 86
 321 GLYLYSerArgAspSerTYRserSerSerArgSerAspLeuTYRser 337
 87GCCTTACAGTCCGAGCA..... 104
 337 rGLYArgAspArgValGLYArgGLNGLYArgLeuProProSerMetG 354
 105AGGGA 110
 354 LuArgGLYTYRProProRoArgAspSerTYRserSerSerArgGLY 370
 111 GATCAAAACCTCGAGAGACAGACGAGTGAAGTGGCGAGTGGAGG 160
 371 AlaProArgGLYGLYGLYArgGLYGLYSerArgSerArgGLYGLY 387
 161 CAGG 164
 387 YArg 388

seq_name: SwissProt_39:RS2_ARATH

seq_documentation_block:
 ID RS2_ARATH STANDARD; PRT; 285 AA.
 AC P49688; O22936;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE 40S RIBOSOMAL PROTEIN S2.
 GN RPS2 OR AT2G41840 OR T11A07.6.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RX MEDLINE=20083487; PubMed=10617197;
 RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D., Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V., Buehl C.R., Ketchum K.A., Lee J.J., Konning C.M., Koo H.L., Moffat K.S., Cronin L.A., Shen M., VanAken S.E., Umayam L., Tallon L.J., Gill J.E., Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R., Copenhaver G.P., Preuss D., Nierman W.C., White O., Eisen J.A., Salzberg S.L., Fraser C.M., Venter J.C.;
 RA "Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.";
 RT Nature 402:761-768(1999).
 RT [2]
 RP SEQUENCE OF 119-285 FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RA Raynal M., Grillet F., Laudie M., Meyer Y., Cooke R., Delserny M.;
 RL Submitted (OCT-1992) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: BELONGS TO THE S5P FAMILY OF RIBOSOMAL PROTEINS.

CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC
 CC EMBL: AC002339; AAC02764.1;
 CC DR EMBL: Z17622; CAA79017.1;

```

DR HSSP; P02357; 1PKP.
DR InterPro: IPR000851; Ribosomal_S5.
DR Pfam: PF00333; Ribosomal_S5_1.
DR PROSITE: PS00585; RIBOSOMAL_S5; 1.
KW Ribosomal protein.
FT CONFLICT 268 269 AS -> ST (IN REF. 2).
FT CONFLICT 272 281 LSTRKDPDV -> VSKRVLTG (IN REF. 2).
SQ SEQUENCE 285 AA; 30878 MW; 68C8C98091DA995B CRC64.

alignment_scores:
    Quality: 66.00      Length: 53
    Ratio: 1.833        Gaps: 3
    Percent Similarity: 67.925      Percent Identity: 37.736

alignment_block:
US-09-823-101-2 x RS2_ARATH ..

Align seg 1/1 to: RS2_ARATH from: 1 to: 285

30 GGGGCTCCAGAGAGCTAGACAGATAGAGTCGAGCTCTCATGCCAA 79
||||| :||| :||| :||| :||| :||| :||| :||| :|||
5 GylgylguarGylgylValGluarGylgylGluarGylgylspheGylar 21
: :||| :||| :||| :||| :||| :||| :||| :||| :|||
80 ATACATAGCGCTTACAGTCCCGAGGAGGAGGAGATCCAAACCCGSAAGAA 129
: :||| :||| :||| :||| :||| :||| :||| :||| :|||
21 gylpheGylgylarGylgylGluarGylgylasp..ArgGylgylarGg 36
: :||| :||| :||| :||| :||| :||| :||| :||| :|||
130 GACAAAGCAGCAGTACAGTACGCGAGTGGCGAGCGAGCGAGG.....AAA 167
: :||| :||| :||| :||| :||| :||| :||| :||| :|||
36 lYarGylgylarGylgylGluarGylgylarGylgylarGylgylalserGluGlu 52
: :||| :||| :||| :||| :||| :||| :||| :||| :|||
168 GGTGGCGGTG 176
|||||
53 lYstrpVal 55

seq_name: SwissProt_39:EMS_HUMAN

seq_documentation_block:
ID EMS_HUMAN STANDARD; PRT; 656 AA.
AC O01844; O92635;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE RNA-BINDING PROTEIN EMS (EMS ONCOGENE) (EWING SARCOMA BREAKPOINT
DE REGION 1 PROTEIN).
GN EMSR1 OR EMS.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
SEQUENCE FROM N.A.
RC TISSUE=Petal brain;
RX MEDLINE=2336239; PubMed=1522903;
RA Delattre O., Zucman J., Plougastel B., Desmazes C., Melot T.,
RA Peter M., Kovar H., Joubert I., de Jong P., Rouleau G., Aurias A.,
RA Thomas G.;
RT "gene fusion with an ETS DNA-binding domain caused by chromosome
RT translocation in human tumours.";
RL Nature 359:162-165(1992).
[2]
SEQUENCE FROM N.A.
RX MEDLINE=94140360; PubMed=8307570;
RA Plougastel B., Zucman J., Peter M., Thomas G., Delattre O.;
RT "genomic structure of the EMS gene and its relationship to EMSR1, a
RT site of tumor-associated chromosome translocation.";
RL Genomics 18:609-615(1993).
[3]
SEQUENCE FROM N.A.
RA Zucman-Rossi J.;
RP Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
[4]

```

RP SEQUENCING OF 1-345 FROM N.A.
RX MEDLINE=97131501; PubMed=8975669;
RA Zucman-Rossi J., Legoux P., Thomas G.;
RT "Identification of new members of the Gas2 and Ras families in the
RT 22q12 chromosome region.";
RL Genomics 38:247-254(1996).
RN 15]
RP SEQUENCING OF 241-268 FROM N.A.
RX TISSUE=Placenta;
RX MEDLINE=95352541; PubMed=7542907;
RA Bhagichand T., Abe S., Nojima T., Yoshida M.C.;
RT "Molecular analysis of a t(11;22) translocation junction in a case of
RT Ewing's sarcoma.";
RL Genes Chromosomes Cancer 13:116-132(1995).
RN 16]
RP ALTERNATIVE SPLICING, AND RNA-BINDING.
RX MEDLINE=9436763; PubMed=8084618;
RA Ohno T., Ouchida M., Lee L., Gatalica Z., Rao V.N., Reddy E.S.P.;
RT "The EWS gene, involved in Ewing family of tumors, malignant melanoma
RT of soft parts and desmoplastic small round cell tumors, codes for an
RL RNA binding protein with novel regulatory domains.";
RN Oncogene 9:3087-3097(1994).
RN 17]
RP CHARACTERIZATION.
RX MEDLINE=20390060; PubMed=10767297;
RA Li K.K.C., Lee K.A.W.;
RT "Transcriptional activation by the Ewing's sarcoma (EWS) oncogene can
RL be cis-repressed by the EWS RNA-binding domain.";
RL J. Biol. Chem. 275:23053-23058(2000).
CC -1- FUNCTION: MIGHT NORMALLY FUNCTION AS A REPRESSOR. 'EWS-
CC FUSION-PROTEINS (EFPs) MAY PLAY A ROLE IN THE TUMORIGENIC PROCESS.
CC THEY MAY DISTURB GENE EXPRESSION BY MIMICKING, OR INTERFERING
CC WITH THE NORMAL FUNCTION OF GYD-POL1 WITHIN THE TRANSCRIPTION
CC INITIATION COMPLEX. THEY MAY ALSO CONTRIBUTE TO AN ABERRANT
CC ACTIVATION OF THE FUSION PROTEIN TARGET GENES.
CC -1- SUBUNIT: BINDS DNA-DIRECTED RNA POLYMERASE II. 33 KDA POLYPEPTIDE
CC (RPB3), RNA AND CALMODULIN. INTERACTS WITH THE FOCAL ADHESION
CC KINASE 2 (FAK2).
CC -1- SUBCELLULAR LOCATION: NUCLEAR AND CYTOPLASMIC. RELOCATES FROM
CC CYTOPLASM TO RIBOSOMES UPON FAK2 ACTIVATION.
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: EWS (SHOWN HERE) AND EWS-B; MAY
CC BE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: UBIDUOUS.
CC -1- DOMAIN: EWS ACTIVATION DOMAIN (EAD) FUNCTIONS AS A POTENT
CC ACTIVATION DOMAIN IN EFPs. EWS BINDS RPB3 BUT NOT RPB5 OR RPB7,
CC WHEREAS THE ISOLATED EAD BINDS RPB5 AND RPB7 BUT NOT RPB3. CIS-
CC LINKED RNA-BINDING DOMAIN (RBD) CAN STRONGLY AND SPECIFICALLY
CC REPRESS TRANS-ACTIVATION BY THE EAD.
CC -1- DISEASE: EWING'S SARCOMA IS CHARACTERIZED BY CHROMOSOMAL
CC TRANSLOCATIONS T(11;22)(Q24;Q12) WHICH INVOLVES EWS AND FLI1,
CC T(7;22)(P22;Q12) WHICH INVOLVES EWS AND ETV1, T(21;22)(Q22;Q12)
CC WHICH INVOLVES EWS AND ERG AND T(9;22)(Q22-31;Q11-12) WHICH
CC INVOLVES EWS AND NR4A3.
CC -1- DISEASE: DESMOPLASTIC SMALL ROUND CELL TUMOR (DSRCT) IS CAUSED BY
CC CHROMOSOMAL TRANSLOCATION T(11;22)(P13;Q12), WHICH INVOLVES EWS
CC AND WT1.
CC -1- DISEASE: MALIGNANT MELANOMA OF SOFT PARTS (MMSF), ALSO KNOWN AS
CC SOFT TISSUE CLEAR CELL SARCOMA, IS A RARE TUMOR DEVELOPING IN
CC TENONS AND APONEUROSES. IT IS ASSOCIATED WITH CHROMOSOMAL
CC TRANSLOCATION T(12;22)(Q13;Q12) INVOLVING EWS AND ATR-1.
CC -1- MISCELLANEOUS: EFPs ARISE DUE TO CHROMOSOMAL TRANSLOCATIONS IN
CC WHICH EWS IS FUSED TO A VARIETY OF CELLULAR TRANSCRIPTION FACTORS.
CC EFPs ARE VERY POTENT TRANSCRIPTIONAL ACTIVATORS DEPENDENT ON THE
CC EAD AND A C-TERMINAL DNA-BINDING DOMAIN CONTRIBUTED BY THE FUSION
CC PARTNER. THE SPECTRUM OF MALIGNANCIES ASSOCIATED WITH EFPs ARE
CC THOUGHT TO ARISE VIA EFP-INDUCED TRANSCRIPTIONAL DEREGLATION,
CC WITH THE TUMOR PHENOTYPE SPECIFIED BY THE EWS FUSION PARTNER AND
CC CELL TYPE. TRANSCRIPTIONAL REPRESSION OF THE TRANSFORMING GROWTH
CC FACTOR BETA TYPE II RECEPTOR (TGF-BETA RII) IS AN IMPORTANT TARGET
CC OF THE EWS-FLI1, EWS-ERG, OR EWS-ETV1 ONCOGENE.
CC -1- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RRM).
CC -1- SIMILARITY: BELONGS TO THE TET FAMILY OF RNP PROTEINS.

DR	EMBL	X66899	CAA47350.1	-	-
DR	EMBL	X72990	CAAS1489.1	JOINED	
DR	EMBL	X72991	CAAS1489.1	JOINED	
DR	EMBL	X72992	CAAS1489.1	JOINED	
DR	EMBL	X72993	CAAS1489.1	JOINED	
DR	EMBL	X72994	CAAS1489.1	JOINED	
DR	EMBL	X72995	CAAS1489.1	JOINED	
DR	EMBL	X72996	CAAS1489.1	JOINED	
DR	EMBL	X72997	CAAS1489.1	JOINED	
DR	EMBL	X72998	CAAS1489.1	JOINED	
DR	EMBL	X72999	CAAS1489.1	JOINED	
DR	EMBL	X73000	CAAS1489.1	JOINED	
DR	EMBL	X73001	CAAS1489.1	JOINED	
DR	EMBL	X73002	CAAS1489.1	JOINED	
DR	EMBL	X73003	CAAS1489.1	JOINED	
DR	EMBL	X73004	CAAS1489.1	JOINED	
DR	EMBL	X08806	CAA70044.1	ALTI_INIT	
DR	EMBL	X08807	CAA70044.1	ALTI_INIT	

SEQUENCE FROM N.A.
TISSUE=Testis;

RA Atkinson M.R., Townsend-Nicholson A., Nicholl J.K., Sutherland G.R.,
 RA Schofield P.R.;
 CC Submitted (JUN-1996) to the EMBL/Genbank/DBJ databases.
 CC -1- FUNCTION: RECEPTOR FOR ADENOSINE. THE ACTIVITY OF THIS RECEPTOR
 CC IS MEDIATED BY G PROTEINS WHICH INHIBITS ADENYLATE CYCLASE.
 CC POSSIBLE ROLE IN REPRODUCTION.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
 CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: L20463; AAA16365.1; -;
 CC EMBL: L22607; AAA35949.1; -;
 CC EMBL: X76981; CAA54288.1; -;
 CC EMBL: L77730; AAB02790.1; -;
 CC EMBL: L77729; AAB02790.1; JOINED.
 CC PIR: S38511; S38511.
 CC HSSP: P29274; IMM.
 CC GCRDB: GCR_0746; -;
 CC GCRDB: GCR_0791; -;
 CC GCRDB: GCR_1878; -;
 CC GCRDB: GCR_2061; -;
 CC MIM: 600445; -;
 CC InterPro: IPR000276; GPCR_Rhodopsn.
 CC Pfam: PF00001; 7tm_1; 1.
 CC PRINTS: PRO0237; GPCR_RHODOPSIN.
 CC PRINTS: PRO0424; ADENOSINER.
 CC PRINTS: PRO0555; ADENOSINER3.
 CC PROSITE: PS00237; G-PROTEIN_RECEPTOR_F1_1; 1.
 CC PROSITE: PS50262; G-PROTEIN_RECEPTOR_F1_2; 1.
 CC DR G-protein coupled receptor; Transmembrane; Glycoprotein;
 CC Lipoic acid; Palmitate.
 CC KM DOMAIN 1 14 EXTRACELLULAR (POTENTIAL).
 CC FT TRANSMEM 15 37 1 (POTENTIAL).
 CC FT DOMAIN 38 48 CYTOPLASMIC (POTENTIAL).
 CC FT TRANSMEM 49 72 2 (POTENTIAL).
 CC FT DOMAIN 73 84 EXTRACELLULAR (POTENTIAL).
 CC FT TRANSMEM 85 106 3 (POTENTIAL).
 CC FT DOMAIN 107 126 CYTOPLASMIC (POTENTIAL).
 CC FT TRANSMEM 127 148 4 (POTENTIAL).
 CC FT DOMAIN 149 177 EXTRACELLULAR (POTENTIAL).
 CC FT TRANSMEM 178 198 5 (POTENTIAL).
 CC FT DOMAIN 199 231 CYTOPLASMIC (POTENTIAL).
 CC FT TRANSMEM 232 255 6 (POTENTIAL).
 CC FT DOMAIN 256 261 EXTRACELLULAR (POTENTIAL).
 CC FT TRANSMEM 262 284 7 (POTENTIAL).
 CC FT DOMAIN 285 318 CYTOPLASMIC (POTENTIAL).
 CC FT DISULFID 83 166 BY SIMILARITY.
 CC FT LIPID 303 303 PALMITATE (POTENTIAL).
 CC FT CARBOHYD 3 3 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC FT CARBOHYD 4 4 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC FT CARBOHYD 12 12 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC FT CONFLICT 7 7 A -> T (IN REF. 1).
 CC FT CONFLICT 248 248 I -> L (IN REF. 3).
 CC SQ SEQUENCE 318 AA; 36184 MW; 690E67986130FC28 CRC64;

alignment_scores:
 Quality: 65.50 Length: 54
 Ratio: 2.113 Gaps: 2
 Percent Similarity: 57.407 Percent Identity: 31.481

alignment_block:
 US-09-823-101-2/rev x AA3R_HUMAN ..

Align seg 1/1 to: AA3R_HUMAN from: 1 to: 318

seq_name: SwissProt_39:FBRL_YEAST
 seq_documentation_block:
 ID FBRL_YEAST STANDARD; PRT; 327 AA.
 AC P15646; P89890;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE FIBRILLARIN (NUCLEOLAR PROTEIN 1).
 GN NOP1 OR YDL014W OR D2870.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC STRAIN=BJ2168;
 RX MEDLINE=90130477; PubMed=2298745;
 RA Henriquez R., Blobel G., Aris J.P.;
 RT "Isolation and sequencing of NOP1. A yeast gene encoding a nucleolar
 RT protein homologous to a human autoimmune antigen.";
 RL J. Biol. Chem. 265:2209-2215(1990).
 RN [2]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC MEDLINE=90076121; PubMed=2686980;
 RX Schlimm T., Tollervey D., Kern H., Frank R., Hurt E.C.;
 RT "A yeast nucleolar protein related to mammalian fibrillarin is
 RT associated with small nucleolar RNA and is essential for viability.";
 RL EMBO J. 8:4015-4024(1989).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288C;
 RA Andre B., Vissers S., Urrestarazu L.;
 RT submitted (FEB-1995) to the EMBL/Genbank/DBJ databases.
 RN [4]
 RP FUNCTION.
 RX MEDLINE=91160511; PubMed=1825809;
 RA Tollervey D., Lenton H., Carmo-Fonseca M., Hurt E.C.;
 RT "The small nucleolar RNP protein NOP1 (fibrillarin) is required for
 RT pre-rRNA processing in yeast.";
 RL EMBO J. 10:573-583(1991).
 CC -1- FUNCTION: FIBRILLARIN IS A COMPONENT OF A NUCLEOLAR SMALL NUCLEAR
 CC RIBONUCLEOPROTEIN PARTICLE THOUGHT TO PARTICIPATE IN THE FIRST
 CC STEP IN PROCESSING PRE-RIBOSOMAL RNA.
 CC -1- SUBUNIT: ASSOCIATES WITH NOP56 AND NOP58.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR; FIBRILLARIN REGION OF THE NUCLEOLUS.
 CC -1- PTM: BY HOMOLOGY TO OTHER FIBRILLARINS, SOME OR ALL OF THE
 CC N-TERMINAL ARGININES ARE N-DIMETHYLATED (DMA).
 CC -1- SIMILARITY: BELONGS TO THE FIBRILLARIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

```

CC -----
DR EMBL: J05230; AAA34816.1; -
DR EMBL: 248432; CAAB8345.1; -
DR EMBL: 274061; CA98571.1; -
DR EMBL: 274062; CA98572.1; -
DR PIR: A35038; A35038.
DR PIR: S25421; S25421.
DR SGD: S0002172; NOP1.
DR InterPro: IPR000692; Fibrillarin.
DR Pfam: PF01269; Fibrillarin; 1.
DR PRINTS: PR00052; FIBRILLARIN.
DR PRODOM: PD004637; Fibrillarin; 1.
DR PROSITE: PS00566; Fibrillarin; 1.
KW Nuclear protein; Methylation; Ribonucleoprotein; rRNA processing;
KW RNA-binding.
FT DOMAIN 6 83 GLY/DNA-RICH.
FT DOMAIN 22 42 RNA-BINDING RGG-BOX (BY SIMILARITY).
FT DOMAIN 58 75 RNA-BINDING RGG-BOX (BY SIMILARITY).
FT DOMAIN 281 313 HELICAL (POTENTIAL).
SQ SEQUENCE 327 AA; 34465 MW; 56AB958A7B6066E CRC64;

```

```

alignment_scores:
Quality: 65.50 Length: 50
Ratio: 2.339 Gaps: 3
Percent Similarity: 56.000 Percent Identity: 40.000

```

```

alignment_block:
US-09-823-101-2 x FBRL_YEAST ..

```

```

Align seg 1/1 to: FBRL_YEAST from: 1 to: 327

```

```

30 GGGGGTCCAGAGAGTAGAGAGATAGAGCTCAGCTCATAGCCAA 79
||||| ::::::::::| | | | | | | | | | | | | | | | | |
41 GlylyserArglyGlyAlaArglyGlySerArgly..... 53
80 ATACTATGCTTACAGTCCGAGAGGAGATCAAACTGGAAGAA 129
:::| | | | | | | | | | | | | | | | | | | | | | |
54 GlyPhegly.....GlyArgglySerArglyGlyAla 66
130 GACAGGACAGTAGAGAGTGGCGA.....GTGGAGGAGGAGAAAGT 170
|| | | | | | | | | | | | | | | | | | | | | | |
66 rglyglySerArglyGlyArgglyGlyAlaGlyAlaGlyAlaGly 82

```

```

seq_name: SwissProt_39:FUS_BOVIN

```

```

seq_documentation_block:

```

```

ID FUS_BOVIN STANDARD; PRT; 512 AA.

```

```

AC Q28009;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE RNA-BINDING PROTEIN FUS (PIGEPEN PROTEIN).
GN FUS.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=orta;
RX MEDLINE=96175600; PubMed=8631501;
RA Allegro M.C.; Allegro M.A.;
RT "A nuclear protein regulated during the transition from active to
RT quiescent phenotype in cultured endothelial cells.";
RL Dev. Biol. 174:288-297(1996).

```

```

RN [2]
RX CARBOHYDRATE BINDING DOMAIN.
RX MEDLINE=20160719; PubMed=10694442;
RA Allegro M.C.;
RT "A C-terminal carbohydrate-binding domain in the endothelial cell
RT regulatory protein, pigpen: new function for an EWS family member.";

```

```

RL Exp. Cell Res. 255:270-277(2000).
CC -1- FUNCTION: BINDS BOTH SINGLE-STRANDED AND DOUBLE-STRANDED DNA AND
CC PROMOTES ATP-INDEPENDENT ANNEALING OF COMPLEMENTARY SINGLE-
CC STRANDED DNAs AND D-LOOP FORMATION IN SUPERHELICAL DOUBLE-STRANDED
CC DNA. MAY PLAY A ROLE IN MAINTENANCE OF GENOMIC INTEGRITY (BY
CC SIMILARITY).
CC -1- SUBUNIT: COMPONENT OF NUCLEAR RIBOPROTEIN COMPLEXES (BY
CC SIMILARITY).
CC -1- SUBCELLULAR LOCATION: NUCLEAR, EXHIBITS DIFFUSE STAINING
CC THROUGHOUT (EXCLUDING NUCLEOLI), TOGETHER WITH A SMALL NUMBER OF
CC INTENSELY STAINED FOCAL POINTS, OR GRANULES, AND PUNCTATE STAINING
CC ALONG THE NUCLEAR ENVELOPE.
CC -1- DOMAIN: THE C-TERMINAL DOMAIN BINDS CARBOHYDRATES.
CC -1- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RRM).
CC -1- SIMILARITY: BELONGS TO THE TET FAMILY OF RNP PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----

```

```

CC EMBL: U26024; AAC13543.1; -
CC HSP: P09651; 1H41.
CC InterPro: IPR000504; RRM.
CC InterPro: IPR001876; Znf-RanBP.
CC Pfam: PF00076; rrm; 1.
CC Pfam: PF00641; zt-RanBP; 1.
CC SMART: SM00360; RRM; 1.
CC SMART: SM00547; Znf_RBZ; 1.
CC PROSITE: PS00102; RRM; 1.
CC PROSITE: PS00030; RRM_RNP_1; FALSE_NEG.
KW RNA-binding; DNA-binding; Nuclear protein; Repeat; Zinc-finger; Zinc;
KW Metal-binding.
FT DOMAIN 1 164 GLN/GLY/SER/TYR-RICH.
FT DOMAIN 165 253 GLY-RICH.
FT DOMAIN 271 357 RNA-BINDING (RRM).
FT DOMAIN 357 512 ARG/GLY-RICH.
FT ZN_FING 414 433 C4-TYPE (POTENTIAL).
SQ SEQUENCE 512 AA; 52240 MW; 3652329C04F1386 CRC64;

```

```

alignment_scores:
Quality: 65.50 Length: 76
Ratio: 1.770 Gaps: 5
Percent Similarity: 48.684 Percent Identity: 32.895

```

```

alignment_block:
US-09-823-101-2 x FUS_BOVIN ..

```

```

Align seg 1/1 to: FUS_BOVIN from: 1 to: 512

```

```

27 AGTGGGGTGTCCAGAGAGTAGAGAGATAGAGCTCAGCTCATAGC 76
||||| | | | | | | | | | | | | | | | | | | | | | |
197 SerGlyGlyArglyGlyGlyGlnGlnAspArgly..... 208
77 CAATACTATGCTTACAGTCCGAGAGGAGATCA..... 116
209 .....GlyArgglyArgglyGlyGlyGlyGlyGlyGlyGly 219
117 .....AACCTGGAGAGACAGACAGAGAGAGT 146
219 snArgSerArglyGlyTyrGlnProArgglyArgglyGlyGlyGly 235
147 GGGGAGTGGAGGACAGAAAGTGGGTGAGACAGAGAGAGTCTCT 196
||||| | | | | | | | | | | | | | | | | | | | | | |
236 GlyArg...GlyGlyMetGlyGly.....SerAspArgglyGly 247
197 GACCTTTTCTTGAGAGAAATCTTAGGC 224
248 .....PheAsnlySphGlyGly 253

```

THIS PAGE BLANK (USPTO)

OM of: US-09-823-101-2 to: SPTREMBL_17.* out_format : pfs

Date: Jan 17, 2002 4:07 PM

About: Results were produced by the Gencore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

```
-MODEL=frimex+n2p.model -DEV=xlp
-O/cqnt2.1/USPTO.spool/US09823101/runat.17012002.155138.22043/app_query.fasta_1.2239
-DB=SPTREMBL_17 -OFMT=fastan -SUFFIX=n2p.rspt -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPEL=0.000 -LOOPEXT=0.000
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FCGAPOP=6.000 -FCGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blomsun62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORWEXT -MINLEN=0 -MAXLEN=200000000
-USER=US09823101@cgn1_1_384 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPPX
-MAIT -THREADS=1
```

Search information block:

```
Query: US-09-823-101-2
Query length: 235
Database: SPTREMBL_17.*
Database sequences: 473505
Database length: 146272329
Search time (sec): 224.230000
```

```
score_list:
Sequence      Strd Orig      ZScore      EScore Len      Documentation
sp_rident:Q9WVC9 - 78.00 142.72 1.29 873 09wvc9 mus musculus (mouse). hy
sp_invertebrate:Q9UAX0 + 75.50 142.18 2.56 471 049490 caenorhabditis elegans
sp_bactera:Q49806 + 75.00 144.60 2.97 298 049806 mycobacterium leprae. b2
sp_invertebrate:Q18265 + 75.00 140.03 2.91 545 018265 caenorhabditis elegans
sp_invertebrate:Q9V456 - 75.00 133.42 2.84 1307 09v456 drosophila melanogast
sp_human:Q94945 - 73.00 136.13 4.98 526 094945 homo sapiens (human). kl
sp_human:Q9HAP2 - 73.00 131.91 4.89 919 09hap2 homo sapiens (human). mc
sp_invertebrate:Q9VU42 - 73.00 128.70 4.83 1406 09v42 drosophila melanogast
sp_plant:Q9AM13 - 72.00 140.17 6.66 234 09am13 oryza sativa (rice). p04
sp_plant:Q41848 - 71.50 137.23 7.56 301 041848 zea mays (maize). p101n
sp_human:Q16824 - 71.50 129.86 7.34 797 016824 homo sapiens (human). fi
sp_fungi:Q94056 + 70.50 136.16 9.91 263 094056 candida albicans (yeast)
sp_invertebrate:Q9GV11 + 70.50 131.44 9.73 491 09gv11 ephydria fluviatilis
sp_invertebrate:Q62213 + 70.00 136.90 11.41 208 062213 caenorhabditis elegans
sp_plant:Q42412 + 70.00 136.86 11.40 209 042412 nicotiana glauca (tobacco)
sp_fungi:Q9HE26 + 70.00 133.14 11.24 342 09he26 neurospora crassa. proba
sp_invertebrate:Q9NKF2 - 69.50 132.01 11.19 221 09nkf2 drosophila melanogast
sp_invertebrate:Q9A407 - 69.50 135.40 13.01 397 09a407 caenorhabditis elegans
sp_human:Q9H413 + 69.50 125.69 12.52 798 09h413 homo sapiens (human). d3
sp_plant:Q9FM5 + 69.50 124.60 12.47 921 09fm5 oryza sativa (rice). put
sp_bactera:Q9XBM1 + 69.00 128.33 14.52 490 09xbm1 prausetella rugosa. alka
sp_rident:Q88539 - 68.50 129.04 16.71 389 088539 mus musculus (mouse). wa
sp_organella:Q9T382 - 68.50 126.95 16.57 513 09t382 eustachys mutica, and ch
sp_human:Q9Y2W2 - 68.50 125.26 16.46 641 09y2w2 homo sapiens (human). sh
sp_invertebrate:Q61920 + 68.50 124.67 16.43 693 061920 caenorhabditis elegans
sp_invertebrate:Q9VPS8 + 68.00 137.76 19.86 107 09vps8 drosophila melanogast
sp_plant:Q9C704 + 68.00 130.43 19.29 282 09c704 arabidopsis thaliana (mc
sp_invertebrate:Q9VRM3 - 68.00 130.32 19.28 286 09vr04 drosophila melanogast
sp_human:Q9H010 - 67.50 132.45 22.32 188 09h010 homo sapiens (human). by
sp_invertebrate:Q9U1A8 + 67.50 132.10 22.29 197 09u1a8 leishmania major. hypc
sp_plant:Q9LGV9 - 67.50 127.65 21.89 355 09lgy9 oryza sativa (rice). hyc
sp_virus:Q9C819 + 67.50 124.72 21.64 523 09c819 arabidopsis thaliana (mc
sp_virus:Q9QCX4 + 67.50 122.73 21.47 680 09qcx4 chalybe mosaic tymovirus
sp_plant:Q9LPI3 + 67.50 119.47 21.34 829 09lpi3 arabidopsis thaliana (mc
sp_invertebrate:Q9VR13 - 67.50 119.47 21.19 1047 09vr13 drosophila melanogast
sp_rident:Q9DB92 + 67.00 132.87 25.66 155 09db92 mus musculus (mouse). 15
sp_plant:Q40425 + 67.00 132.40 25.61 165 040425 nicotiana glauca (tobacco)
sp_invertebrate:Q9N9T4 - 67.00 130.43 25.41 214 09n9t4 leishmania major. hypc
sp_plant:Q9S221 - 67.00 127.39 25.10 320 09s221 arabidopsis thaliana (mc
sp_rident:Q62775 - 67.00 124.79 24.84 451 062775 rattus norvegicus (rat).
```

```
seq_name: sp_rident:Q9WVC9
seq_documentation_block:
ID 09WVC9 PRELIMINARY; PRT: 873 AA.
AC 09WVC9;
DT 01-NOV-1999 (TREMREL. 12. Created)
DT 01-NOV-1999 (TREMREL. 12. Last sequence update)
DT 01-JUN-2001 (TREMREL. 17. Last annotation update)
DE HUNTINGTON YEAST PARTNER C.
GN 2610317D23RIK OR HYPC.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Bedford M.T., Das R., Reed R., Leder P.
RT "EPPII, a mammalian ortholog of the essential yeast splicing factor
RT PRP40."
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF135440; AAD39464.1;
DR MGD: MGI:1925583; 2610317D23RIK.
DR InterPro: IPR002713; FF.
DR InterPro: IPR002965; P-rich_extensn.
DR InterPro: IPR001202; WW.
DR Pfam: PF01846; EF; 5.
DR Pfam: PF00397; WW; 2.
DR PRINTS: PR01217; PRICHEXTENS.
DR SMART: SM00441; FF; 4.
DR SMART: SM00456; WW; 2.
DR PROSITE: PS01159; WW_DOMAIN_1; 1.
DR PROSITE: PS00202; WW_DOMAIN_2; 2.
SQ SEQUENCE 873 AA; 99697 MW; E3FCFDBC3725A5FE CRC64;
```

alignment_scores:
Quality: 78.00 Length: 55
Ratio: 2.889 Gaps: 1
Percent Similarity: 49.091 Percent Identity: 32.727

alignment_block:
US-09-823-101-2/rev x Q9WVC9 ..

Align seg 1/1 to: Q9WVC9 from: 1 to: 873

```
190 CCGCCGTGATGCTGACCAACCTTCGCGCCGACCTGCGCCACTTCG 141
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
22 PPGPMPGSetelnaPProProalalleProPrometProolylt 38
140 ACTGCGTTCCTTCCTTCAGAGGTTTGATGATCCCTTCCTCGGAGCTGA 91
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
38 elenPProPrometleuProPrometcllyalProProPro..... 51
90 AGCCATGATGATTTGGCTATGAGAGCCTGACTTATCTCTCTACTCTCT 41
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
52 .....leuThrInleProgly 57
40 TGGACCGCCGCACTC 26
58 MetValProPromet 62
seq_name: sp_invertebrate:Q9UAY0
seq_documentation_block:
ID 09UAY0 PRELIMINARY; PRT: 471 AA.
AC 09UAY0;
DT 01-MAY-2000 (TREMREL. 13. Created)
```

DT 01-MAY-2000 (TREMBlrel. 13, last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, last annotation update)
 DE W03G1.5 PROTEIN.
 GN W03G1.5.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Pelodierinae; Caenorhabditis.
 OX NCBI_TaxID=6239.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BRISTOL N2;
 RX MEDLINE=94150718; PubMed=7906398;
 RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
 RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
 RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
 RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
 RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
 RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
 RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkneen R.,
 RA Smaildon N., Smith A., Sonhammer E., Staden K., Sulston J.,
 RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
 RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.;
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 RT elegans.";
 RL Nature 368:32-38(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BRISTOL N2;
 RA Pauley A., Scheet P., Harper M.;
 RT "The sequence of C. elegans cosmid W03G1.";
 RT Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BRISTOL N2;
 RA Waterston R.;
 RA Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF125964; AAD14753.1; -; BDF30B59A64A9B58 CRC64;
 SO SEQUENCE 471 AA; 50885 MW; BDF30B59A64A9B58 CRC64;

alignment_scores:
 Quality: 75.50 Length: 49
 Ratio: 2.517 Gaps: 2
 Percent Similarity: 61.224 Percent Identity: 42.857

alignment_block:
 US-09-823-101-2 x Q9UAY0 ..
 Align seg 1/1 to: Q9UAY0 from: 1 to: 471

seq_documentation_block:
 ID 049806 PRELIMINARY; PRT: 298 AA.
 AC 049806;
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-NOV-1996 (TREMBlrel. 01, Last annotation update)
 DE B2126.F3.142.
 OS Mycobacterium leprae.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;

OC Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1769;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Smith D.R.;
 RL Submitted (JAN-1994) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Robison K.;
 RL Submitted (NOV-1993) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC Robison K.;
 RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
 DR EMBL: U00017; AAA17220.1; -;
 SO SEQUENCE 298 AA; 30858 MW; B3573B073E2139C3 CRC64;

alignment_scores:
 Quality: 75.00 Length: 55
 Ratio: 2.143 Gaps: 1
 Percent Similarity: 63.636 Percent Identity: 34.545

alignment_block:
 US-09-823-101-2/rev x Q49806 ..
 Align seg 1/1 to: Q49806 from: 1 to: 298

185 CTGATCTGTACCCACCTTCCGCTCCGACCTGCGCCACTTCTACTGC 136
 ||||| : : : ||||| : : : ||||| : : : ||
 212 LeuileuAarglnThrTyProValArgHisAlaProTyrglyLeuAl 228
 135 CTGCTCTTCTCCAGGTTTGATCTCCCTCCGCGGACTGTAAACA 86
 | || : : : ||||| : : : ||||| : : : ||
 228 aglySerArghrleuAlatrpValserGlnThrAlser.....AsnI 243
 85 TAGATATTTGGCTATGAGAGCCCTGACTTATCTCTTCTTACCTCTTGAC 36
 ||||| : : : ||||| : : : ||||| : : : ||
 243 levalileglyTyraSplleuThrThrGlyleProValGluIysValser 259
 35 ACCCCCACTCACCA 21
 ||||| |||
 260 TyreProThrValGln 264

seq_name: sp_invertebrate:Q18265

seq_documentation_block:
 ID Q18265 PRELIMINARY; PRT: 545 AA.
 AC Q18265;
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE SIMILAR TO NUCLEOLIN.
 GN C27H5.3.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Pelodierinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BRISTOL N2;
 RX MEDLINE=94150718; PubMed=7906398;
 RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
 RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
 RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
 RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
 RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
 RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
 RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkneen R.,
 RA Smaildon N., Smith A., Sonhammer E., Staden K., Sulston J.,
 RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
 RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.;
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.

RT elegans." ;
 RL Nature 368:32-38(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA Pauley A.;
 RL Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA Waterston R.;
 RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.
 DR EMBL; U14635; AAC4657.1; .
 DR InterPro: IPR000504; RRM.
 DR InterPro: IPR001876; ZnF-RanBP.
 DR Pfam: PF00076; rrm; 1.
 DR Pfam: PF00641; zf-RanBP; 1.
 DR PROSITE; PS0102; RRM; 1.
 DR SMART; SM00360; RRM; 1.
 DR SMART; SM00547; ZnF_RB2; 1.
 SQ SEQUENCE 545 AA; 56896 MW; 36D771CDA39A0CAE CRC64;

alignment_scores:
 Quality: 75.00 Length: 65
 Ratio: 2.419 Gaps: 3
 Percent Similarity: 47.692 Percent Identity: 38.462

alignment_block:
 US-09-823-101-2 * Q18265 ..

Align seg 1/1 to: Q18265 from: 1 to: 545

```

27 AGTGGGGGTCCTCAAGAGTGAAGAGATAGAGTCAGCTCTCATAGC 76
   |||||||  ::|||
123 SeeglyglGlnserglYglY.....Se 130
   77 CAAATCTAGTGGCTTACAGTCCCGAGAGGAGATCCAAAC..... 119
   :      |||||  |||||||  ||
130 rasprotYrgly..GlnserarglyglYarglyglYPhcglyg 146
   120 .....CCTGAAGAAGACAGAGGCACTAGAGT 146
   146 lyserarglyglYglYTYraspglyglYarglyglYserargly 162
   147 GGGCGAGTGGAGCGCAGAAAGTTGGTGACAGATCAGGAGG 191
   ||      |||||||  |||::  |||::|||
163 glyTYraspglyglYarglyglYTYrglyglYAsparglyglY 177

```

seq_name: sp_Invertebrate:Q9V4J6

seq_documentation_block:

```

ID Q9V4J6      PRELIMINARY:      PRT: 1307 AA.
AC Q9V4J6;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE CG11101 PROTEIN.
CN CG11101.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BERKELEY;
RA MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Landell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,

```

```

RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Bailew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brockstein P., Brotlier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferrara C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon R., Nussken D.R., Pauleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodard T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RA "The genome sequence of Drosophila melanogaster." ;
RL Science 287:2185-2195(2000).
DR EMBL; AE003841; AAF59274.2; .
DR HSP; P35555; IEMN.
DR FlyBase; FBgn003157; CG11101.
DR InterPro: IPR000152; Asx_hydroxyl.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR001881; EGF-Ca.
DR InterPro: IPR002965; P-rich_extensn.
DR Pfam; PF00008; EGF; 2.
DR PRINTS; PRO1217; PRICHEXTENS.
DR SMART; SM00179; EGF_CA; 1.
DR PROSITE; PS00010; ASX_HYDROXYL; 1.
DR PROSITE; PS00022; EGF_1; UNKNOWN_1.
DR PROSITE; PS01186; EGF_2; UNKNOWN_1.
DR PROSITE; PS01187; EGF_CA; 1.
KW Calcium-binding; EGF-like domain; Glycoprotein; Hydroxylation; Repeat.
SQ SEQUENCE 1307 AA; 139890 MW; 42827878C1B77510 CRC64;

```

alignment_scores:
 Quality: 75.00 Length: 85
 Ratio: 1.744 Gaps: 5
 Percent Similarity: 50.588 Percent Identity: 29.412

alignment_block:
 US-09-823-101-2/rev x Q9V4J6 ..

Align seg 1/1 to: Q9V4J6 from: 1 to: 1307

```

229 CACTTCCCTAAGAAATTCTCTCAAGAAAG.....GTCAAGACA 192
   ::|||  ::|||  ::|||
235 AsnleuprolleasptyrGluAlaGlnLysProGlyThrsenleargpr 251
   191 CCTCCCTGATCTGTGTCACCCACCTTCTCGCTCCCACTCGCCCACTTC 142
   |||||  |||||  |||||
251 opfropo.....TyrProheasparpArgGluAtrpPhep 264
   141 TACTGCGTGTCTTCTTCACAGGTTTGATCTCCCTCTCTCGGAGCTGT 92
   ::      |||  |||::|||::|||
264 lerhr.....GlyValProlleProGluGlnIle 273

```

```

91 AACCATAGTATTGGCTATGAGAGCCGTACTTATCTCTGCTACCTCC 42
   |||
   :|:|:|
274 ValPro.....ValalProGlypheValSerValAsnArgProPr 286
41 TTGG.....ACACCCCACTGACCAAGTACGACGACTGGG 7
   |||
286 OTTPYrArgSnlYsProAlrProProIleAlrSerValGlyGlyA 303
6 AGAGG 2
   :|:|:|
303 snArg 304

```

seq_name: sp_human:094945

```

seq_documentation_block:
ID 094945 PRELIMINARY; PRT; 526 AA.
AC 094945;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE KIAA0867 PROTEIN.
GN KIAA0867.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE=99156230; PubMed=10048485;
RA Nagase T., Ishikawa K., Suyama M., Kikuno R., Hirosewa M.,
RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XII.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 5:355-364 (1998).
DR EMBL: AB020674; BAA74890.1; -
DR InterPro: IPR001092; HLH_dim.
DR Pfam: PF00010; HLH; 1.
DR PRINTS: PRO1217; PRICHEXTENSIN.
DR SMART: SM00353; HLH; 1.
SQ SEQUENCE 526 AA; 56822 MW; 0C3ADB0852823B60 CRC64;

```

alignment_scores:

Quality:	73.00	Length:	51
Ratio:	2.920	Gaps:	3
Percent Similarity:	49.020	Percent Identity:	43.137

alignment_block:

US-09-823-101-2/rev x 094945 ..

Align seg 1/1 to: 094945 from: 1 to: 526

```

190 CTTCCCTGATCTGTACCCCAACTTCTGCTGCC...ACTGCCCACT 144
   |||
   :|:|:|
27 ProProLeuSerValProGlnProPheLeuProValPheIleMetProLe 43
143 TCTACTGCTTGTCTTCTCCAGGGTTGGATCTGCCCTTCCTCGGGACT 94
   |||
   :|:|:|
43 uLeuSerPro.....SerProAlaProProProI 53
GN
93 GTAAGCATAGTATTGGCTATGAGAGCCGTACTTATCTCTGCTACCT 44
   :|:|:|
53 leSerPro.....ValleuProLeuValPro 61
43 CCT 41
   |||
62 Pro 62

```

seq_name: sp_human:09HAP2

```

seq_documentation_block:
ID 09HAP2 PRELIMINARY; PRT; 919 AA.
AC 09HAP2;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE MONDOA.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20528627; PubMed=11073985;
RA Billin A.N., Eilers A.L., Coulter K.L., Logan J.S., Ayer D.E.;
RT "Monda, a Novel Basic Helix-Loop-Helix-Leucine Zipper Transcriptional
RT Activator That Constitutes a Positive Branch of a Max-Like Network.";
RL Mol. Cell. Biol. 20:8845-8854 (2000).
DR EMBL: AF312918; AAG34121.1; -
DR InterPro: IPR001899; Gram_pos_anchor.
DR Pfam: PF00010; HLH; 1.
DR SMART: SM00353; HLH; 1.
DR PROSITE: PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
SQ SEQUENCE 919 AA; 101113 MW; 47375FC8B6CAAC2 CRC64;

```

alignment_scores:

Quality:	73.00	Length:	51
Ratio:	2.920	Gaps:	3
Percent Similarity:	49.020	Percent Identity:	43.137

alignment_block:

US-09-823-101-2/rev x 09HAP2 ..

Align seg 1/1 to: 09HAP2 from: 1 to: 919

```

190 CTTCCCTGATCTGTACCCCAACTTCTGCTGCC...ACTGCCCACT 144
   |||
   :|:|:|
420 ProProLeuSerValProGlnProPheLeuProValPheIleMetProLe 436
143 TCTACTGCTTGTCTTCTCCAGGGTTGGATCTGCCCTTCCTCGGGACT 94
   |||
   :|:|:|
436 uLeuSerPro.....SerProAlaProProProI 446
GN
93 GTAAGCATAGTATTGGCTATGAGAGCCGTACTTATCTCTGCTACCT 44
   :|:|:|
446 leSerPro.....ValleuProLeuValPro 454
43 CCT 41
   |||
455 Pro 455

```

seq_name: sp_invertebrate:09YU42

```

seq_documentation_block:
ID 09YU42 PRELIMINARY; PRT; 1406 AA.
AC 09YU42;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE CG10600 PROTEIN.
GN CG10600.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,

```

RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazek R.G., Champe M., Pfeiffer B.D.,
 RA Man K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abtill J.F., Agbayan A., An H.-J., Andrews Pfinkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borikova D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,
 RA Butts K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pallos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Domes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasse K.,
 RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idegawa C.,
 RA Jallat M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Laako P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paciel J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert R., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spler E., Spreading A.C., Stapleton M., Strong R., Sun E.,
 RA Styksas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissensbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
 RT "the genome sequence of *Drosophila melanogaster*."
 RL Science 287:2185-2195(2000).
 DR EMBL: AE003660; AAF53715.2; -
 DR FlyBase: FBgn0032717; CG10600.
 DR InterPro: IPR002965; P-rich extensn.
 DR PRINTS: PR01217; PRICHEXTNSN.
 SQ SEQUENCE 1406 AA; 153642 MW; 928070CF036CF4D3 CRC64;

alignment_scores:
 Quality: 73.00 Length: 70
 Ratio: 2.281 Gaps: 2
 Percent Similarity: 45.714 Percent Identity: 30.000

alignment_block:

US-09-823-101-2/rev x Q9VJ42 ..

Align seg 1/1 to: Q9VJ42 from: 1 to: 1406

226 TTGCCTAGAAATTTCTCTCAAGAAAAAGTCAGACACCTCCCTGATCTGT 177
 :::::::::::::::::::::
 1245 ILeProLysSerPheLeu.....HisProProValProI 1256
 :::::::::::::::::::::
 176 CACCCAACTTTCTGCGCTCCACGCTGCCACTGTACAGCGCTGTCTTC 127
 :::::::::::::::::::::
 1256 eValAlaAlaProAlaProProProProSerInIleValLysValL 1273
 :::::::::::::::::::::
 1273 euProGlySerLeuThrProLeuThrArg..... 1282
 :::::::::::::::::::::
 76 GCTATGAGAGCTGACTTATCTCTCTCTGACCTTGAGACACCCCACT 27
 :::::::::::::::::::::
 1283MetSerValAlaLeuProProLeuProProLe 1293
 :::::::::::::::::::::
 26 CACCAAGTCA 17
 |||
 1293 uProProSer 1296

seq_name: sp_plant:Q9AWI3
 seq_documentation_block:
 ID Q9AWI3 PRELIMINARY; PRT: 234 AA.
 AC Q9AWI3;
 DT 01-JUN-2001 (TREMBlrel. 17, Created)
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE P0489A05.17 PROTEIN.
 GN P0489A05.17.
 OS Oryza sativa (Rice).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Eihartoideae; Oryzeae; Oryza.
 ON NCBI_TaxID=4530;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV; NIPPONBARE;
 RA Sasaki T., Matsumoto T., Yamamoto K.;
 RT "Oryza sativa nippobare(GA3) genomic DNA, chromosome 1, PAC
 RT clone:P0489A05.";
 RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AP003105; BAB32994.1; -
 SQ SEQUENCE 234 AA; 24551 MW; ABCDB8DB100D26E CRC64;

alignment_scores:
 Quality: 72.00 Length: 49
 Ratio: 2.323 Gaps: 1
 Percent Similarity: 63.265 Percent Identity: 34.694

alignment_block:

US-09-823-101-2 x Q9AWI3 ..

Align seg 1/1 to: Q9AWI3 from: 1 to: 234

45 GGTAGGAGATAGATAGAGTCAGCTCTCATAGCCAAATACCTAGCTTACA 94
 ||||||| :: ||||| :::
 155 GLYArgArgCysSerArgProAlaValAlaThrThrArgSerAlaAla 171
 :::::::::::::::::::::
 95 GTCCCGAGAGAGGAGATCCAAACCTCGAGAGACAGACAGCATAGAA 144
 :::::::::::::::::::::
 171 agInArgGlyGluGlyGluGluArgGluGlyLys.....GlyGluArgL 186
 ::||| ||| ||||||||| :::::::::::
 145 GTGGCGAGCTGGAGGAGCAGAAAGTTGGTGACAGATCAGCAGAGG 191
 ::||| ||| ||||||||| :::::::::::
 186 yGclYwctGluGluArgLysGlyTTPArgSerAsnGluArgGly 201
 seq_name: sp_plant:Q41848
 seq_documentation_block:
 ID Q41848 PRELIMINARY; PRT: 301 AA.
 AC Q41848;
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE PROLIN RICH PROTEIN.
 GN PRP.
 OS Zea mays (Maize).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
 OC Panicoidae; Andropogoneae; Zea.
 ON NCBI_TaxID=4577;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=W64A;
 RX MEDLINE=92361259; PubMed=1498600;
 RA Jose-Estanyol M., Ruiz-Avila L., Puigdomenech P.;
 RT "A maize embryo-specific gene encodes a proline-rich and hydrophobic
 protein.";
 RL Plant Cell 4:413-423(1992).
 DR EMBL: X60432; CAA42959.1; -.


```

33 GGTGTCCAGAGAGTAGAGATAGACTCAGCTCTCATGACCAATA 82
    ||| :||| ||| :||| :||| :||| :||| :||| :||| :|||
146 GtlyglyArgglyTyrArgspglyArg.....GtlyglyTyr 159
    ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
83 CTAATGCTTACAGTCCCGAGAGAGGAGATCCAAACCTGAGAGAGC 132
    ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
159 rGlyglyTyrArgspglyArgglyTyrGlyglyTyrArgA 176
    ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
133 AAGCAGTAGAGAGTGG..... 149
    ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
176 spGlyglyArgglyTyrArgspglyTyrArgspglyTyr 192
    ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
150 ...CGAGTGGAGCAGCAAGAGCTGGTGCACATCAGGAGG 191
    ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
193 TyrArgspglyArgglyTyrGlyglyTyrGlyglyArggly 207
    ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

```

seq_name: sp_Invertebrate:Q9GV11

```

seq_documentation_block:
ID   Q9GV11      PRELIMINARY;   PRT;   491 AA.
AC   Q9GV11;
DT   01-MAR-2001 (TREMBlrel. 16, Created)
DT   01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DE   01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE   PL10-RELATED PROTEIN POP10 (FRAGMENT).
GN   POP10.
OS   Ephydatia fluviatilis.
OC   Eukaryota; Metazoa; Porifera; Demospongiae; Ceractinomorpha;
OC   Haplosclerida; Spongillidae; Ephydatia.
OX   NCBI_TaxID=31330;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Mochizuki K., Fujisawa T.;
RT   "PL10-related gene in sponge.";
RL   Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
CC   -1- SIMILARITY: TO DEAD/DEAH BOX HELICASE FAMILY.
CC   -1- SIMILARITY: TO HELICASE C-TERMINAL DOMAIN.
DR   EMBL; AB047384; BAB1309.1; -.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR001650; Helicase_C.
DR   Pfam; PF00270; DEAD. 1.
DR   Pfam; PF00271; Helicase_C. 1.
DR   SMART; SM00487; DEXDC; 1.
DR   SMART; SM00490; HELICC; 1.
DR   Kyr ATP-binding; Helicase.
FT   NON_TER 1
SO   SEQUENCE 491 AA; 54545 MW; C77CACB33B642FE3 CRC64;

```

alignment_scores:

Quality:	70.50	Length:	60
Ratio:	2.518	Gaps:	1
Percent Similarity:	46.667	Percent Identity:	36.667

alignment_block:

US-09-823-101-2 x Q9GV11 ..

Align seg 1/1 to: Q9GV11 from: 1 to: 491

```

27 AGTGGGCTGTCACAGAGAGTAGAGATTAAGATCAGGCTCTCATAGC 76
    ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
441 SerGlyglyArgglyTyrArgspglyTyrArg..... 451
    ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
77 CAATTAATCTATGCTTACAGTCCCGAGAGAGGAGATCCAAACCTGGAA 126
    ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
452 .....GlnSerArgglyTyrGlyglyTyrGlyglyGlnSerPheA 463
    ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
127 GAAGACAGCAGTAGAGAGTGGCGAGTGGAGCGAGCAAGAGTTGGGTG 176
    ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
463 spArgglyTyrGlyglyArgglyTyrGlyglyArgglyTyrGlyglyMetSer 479
    ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
177 ACAGATCAGGAGAGGTGTCTGACCTTTTTC 206
    ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

```

480 GtlyglyGlyGlyGlySerValaspTirp 489

seq_name: sp_Invertebrate:062213

seq_documentation_block:

```

ID   062213      PRELIMINARY;   PRT;   208 AA.
AC   062213;
DT   01-AUG-1998 (TREMBlrel. 07, Created)
DT   01-AUG-1998 (TREMBlrel. 07, Last sequence update)
DT   01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE   F33A8.3 PROTEIN.
GN   F33A8.3.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Matthews L.;
RL   Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94150718; PubMed=7906398;
RA   Wilson R., Ainscough R., Anderson K., Baynes C., Berts M.,
RA   Bonfield J., Burton J., Connell M., Copey T., Cooper J., Coulson A.,
RA   Craxton M., Dear S., Du Z., Durbin R., Faveilo A., Fulton L.,
RA   Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA   Jones M., Kershaw J., Kistner J., Laister N., Latreille P.,
RA   Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA   Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,
RA   Smaildon N., Smith A., Sonhammer E., Staden R., Sulston J.,
RA   Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA   Watson A., Weinstock L., Wilkinson-Sproat J., Woldman P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   Nature 368:32-38(1994).
CC   -1- SIMILARITY: BELONGS TO THE COLD-SHOCK DOMAIN (CSD) FAMILY.
DR   EMBL; Z81525; CAB04257.1; -.
DR   HSSP; P41016; IC90.
DR   InterPro; IPR002059; Cold_shock.
DR   Pfam; PF00313; CSD; 1.
DR   PRINTS; PR00050; COLDSHOCK.
DR   ProDom; PD000621; Cold_shock; 1.
DR   SMART; SM00357; CSP; 1.
DR   PROSITE; PS00352; COLD_SHOCK; 1.
KW   DNA-binding; Transcription regulation.
SO   SEQUENCE 208 AA; 21349 MW; 2A6C81AAE3D85FC8 CRC64;

```

alignment_scores:

Quality:	70.00	Length:	77
Ratio:	1.892	Gaps:	4
Percent Similarity:	48.052	Percent Identity:	35.065

alignment_block:

US-09-823-101-2 x 062213 ..

Align seg 1/1 to: 062213 from: 1 to: 208

```

30 GGGGCTGTCACAGAGAGTAGC.....ACAGATTAAGAGTAGGC 67
    ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
93 GlyProValGlnGlySerIlySTYrAlaAlaAspArgspAlaGluVal 109
    ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
68 TCTCATAGCAATACTATGAGCTTACAGTCCGAGAGAG..... 107
    ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
109 a.....AlaArgglyArgglyTyrGlyglyArgglyTyrArgglyTyrA 124
    ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
108 .....GGA 110
    ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
124 rGlyglyTyrGlyglySerArgspArgspAlaGluGluGlygly 140
    ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
111 GATCCAAACCTGGAGAGAGCAGCAGTACAGAGTGGCGAGTGGAGG 160
    ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

```

141 AlaProArgGlyGlyArgGlyGlySerArgArgGlyGly... Gly1 156
 161 CAGGAAGGTTGGGTGACAGATCAGGAGG 191
 |||||
 156 YArgGlyGlyGlyArgGlyThrAsnSerGlyGly 166
 seq_name: sp_Plant:Q42412

seq_documentation_block:
 ID Q42412 PRELIMINARY; PRT; 209 AA.
 AC Q42412;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE RNA-BINDING PROTEIN R2-1.
 OS Nicotiana sylvestris (Wood tobacco).
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Asteridae; euasterids I; Solanales; Solanaceae; Nicotiana.
 OX NCBI_TaxId=4096;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96397973; Pubmed=8804857;
 RA Hanano S., Sugita M., Sugitara M.;
 RT "Structure and expression of the tobacco nuclear gene encoding RNA-
 RT binding protein R2-1: the existence of an intron in the 3'-
 RT untranslated region."
 RL DNA Res. 3:65-71(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=YOUNG LEAF;
 RA Sugitara M.;
 RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE=YOUNG LEAF;
 RA Hanano S., Sugita M., Sugitara M.;
 RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: BELONGS TO ZN-FINGER CCHC TYPE FAMILY.
 DR EMBL; D83696; BAA12064.1; -.
 DR EMBL; D28861; BAA06012.1; -.
 DR Mendel; 15272; Nicsy:2406;15272.
 DR InterPro; IPR000504; RRM.
 DR InterPro; IPR001878; Znf_CCHC.
 DR Pfam; PF00076; xrm; 1.
 DR Pfam; PF00098; zf-CCHC; 1.
 DR PRINTS; PR00939; C2HCZNFINGER.
 DR SMART; SM00360; RRM; 1.
 DR SMART; SM00343; ZNF_C2HC; 1.
 DR PROSITE; PS50102; RRM; 1.
 DR PROSITE; PS00030; RRM_RNP_1; 1.
 DR RNA-binding; Zinc-finger.
 SQ SEQUENCE 209 AA; 22457 MW; 72BBF5EA263569F5 CRC64;

alignment_scores:
 Quality: 70.00 Length: 48
 Ratio: 2.692 Gaps: 1
 Percent Similarity: 54.167 Percent Identity: 41.667

alignment_block:
 US-09-823-101-2 x Q42412 ..

Align seg 1/1 to: Q42412 from: 1 to: 209

27 AGTGGGGCTCCAGAGAGTAGAGAGATAGAGTCAGGCTCATAGC 76
 ||| ||| ::||| |||||
 144 SerGlnGlyGlyArgGlyArgGlyGlyGlyGlyGlyGly..... 157
 77 CAATACTAGGCTTACAGTCCGAGAGAGGAGATCCAAACCCGTGAA 126
 ||| ::||| ||| |||
 158GlySerArgSerSerGlyTyGlyProAspArgAsnGlyA 171

127 GAAGCAAGCGAGTAGAGTGGCGAGTGGAGCGAGGAGGAAAGCT 170
 ||| ||||| ||||| |||||
 171 spArgTyGlySerArgSerGlyArgAspGlyGlyGlyArgGly 185

THIS PAGE BLANK (USPTO)

CDS join(264..474,605..699)
 FT /product='human gastrin'
 FT sig_peptide 264..437
 FT mat_peptide join(438..474,605..699)
 FT /product='human gastrin mature peptide' FT
 3'UTR 700..831.
 FEATURES
 source Location/Qualifiers
 1..829
 /organism='Homo sapiens'
 /db_xref='taxon:9606'
 BASE COUNT 199 a 250 c 213 g 167 t
 ORIGIN
 Query Match 45.3%; Score 47.6; DB 6; Length 829;
 Best Local Similarity 92.4%; Pred. No. 1.2e-05;
 Matches 61; Conservative 0; Mismatches 4; Indels 1; Gaps 1;
 40 gatgcagcagatgtgtatgtctctcttcacatagacagccgctctctg 99
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 261 GATGCAGGACTATGTGTATGTCTGATCTTGCACT-GGCTCTGGCCGCTTCTCTG 319
 QY 100 aagctt 105
 |||||
 Db 320 AACCTT 325
 RESULT 4
 HSGAS1 829 bp DNA PRI 10-FEB-1999
 LOCUS Human gastrin gene.
 DEFINITION X00183
 ACCESSION X00183.1 GI:31648
 VERSION X00183.1
 KEYWORDS gastrin.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE
 1 (bases 1 to 829)
 Kato,K., Hayashizaki,Y., Takahashi,Y., Himeno,S. and Matsubara,K.
 TITLE Molecular cloning of the human gastrin gene
 JOURNAL Nucleic Acids Res. 11 (23), 8197-8203 (1983)
 MEDLINE 84169471
 FEATURES
 source Location/Qualifiers
 1..829
 /organism='Homo sapiens'
 /db_xref='taxon:9606'
 34..42
 /note='variant of CAAT-box'
 64..69
 /note='Goldberg-Hogness box'
 151..472
 /note='exon 1 (472 is 1st base in codon)'
 151..796
 /note='putative primary transcript'
 join(262..472,603..697)
 /note='precursor'
 /codon_start=1
 /protein_id='CAA25005.1'
 /db_xref='GI:31649'
 /db_xref='SWISS-PROT:P01350'
 /translation='MORLCVYLIFALALAFSEASWPKPSQCPDADLTGANDRLT
 PWLEQGPASHHRQLPQGPQPHLVADPSPKQGWLEEEEAAYGMDFGRSADEN'
 262..318
 join(436..472,603..667)
 /note='Protein sequence is in conflict with the conceptual
 translation'
 /codon_start=1
 /product='G34 (big gastrin)'
 /protein_id='CAA25006.1'
 /db_xref='GI:1335072'
 /db_xref='SWISS-PROT:P01350'
 sig_peptide
 CDS

Intron
 mRNA
 CDS
 /translation='QLGPGPHLVADPSKKQGPWLEEEEAAYGWMDF'
 473..602
 /note='intron'
 603..796
 /note='exon 2 (603 is 2nd base in codon)'
 617..667
 /note='Protein sequence is in conflict with the conceptual
 translation'
 /codon_start=1
 /product='G17 (little gastrin) with full biological
 activity'
 /protein_id='CAA25007.1'
 /db_xref='GI:1335073'
 /db_xref='SWISS-PROT:P01350'
 /translation='QGPWLEEEEAAYGWMDF'
 793
 /note='putative polyadenylation site'
 797
 /note='putative polyadenylation site'
 polyA_site
 polyA_site
 polyA_site
 BASE COUNT 198 a 251 c 213 g 167 t
 ORIGIN
 Query Match 45.3%; Score 47.6; DB 9; Length 829;
 Best Local Similarity 92.4%; Pred. No. 1.2e-05;
 Matches 61; Conservative 0; Mismatches 4; Indels 1; Gaps 1;
 40 gatgcagcagatgtgtatgtctctcttcacatagacagccgctctctg 99
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 Db 261 GATGCAGGACTATGTGTATGTCTGATCTTGCACT-GGCTCTGGCCGCTTCTCTG 319
 QY 100 aagctt 105
 |||||
 Db 320 AACCTT 325
 RESULT 5
 HUMGAST2 1217 bp DNA PRI 08-AUG-1995
 LOCUS Human gastrin gene, complete cds.
 DEFINITION K01254 J00147
 ACCESSION K01254.1 GI:162987
 VERSION K01254.1
 KEYWORDS Alu repeat; gastrin; repeat region.
 2 of 2
 SOURCE Human fetal liver genomic DNA (Charon 4A library of Lawn et al.),
 and gastrinoma, cDNA to mRNA (see comment).
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE
 1 (bases 601 to 815; 946 to 1136)
 Kato,K., Himeno,S., Takahashi,Y., Wakabayashi,T., Tarui,S. and
 Matsubara,K.
 TITLE Molecular cloning of human gastrin precursor cDNA
 JOURNAL Gene 26 (1), 53-57 (1983)
 MEDLINE 84159488
 REFERENCE
 2 (bases 600 to 815; 946 to 1139)
 Boel,E., Vuust,J., Norris,F., Norris,K., Wind,A., Rehfeld,J.F. and
 Marcker,K.A.
 TITLE Molecular cloning of human gastrin cDNA: evidence for evolution of
 gastrin by gene duplication
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 80 (10), 2866-2869 (1983)
 MEDLINE 83221503
 REFERENCE
 3 (bases 471 to 1151)
 Wiborg,O., Berglund,L., Boel,E., Norris,F., Norris,K.,
 Rehfeld,J.F., Marcker,K.A. and Vuust,J.
 TITLE Structure of a human gastrin gene
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 81 (4), 1067-1069 (1984)
 MEDLINE 84144842
 REFERENCE
 4 (bases 1 to 1217)
 Ito,R., Sato,K., Helmer,T., Jay,G. and Agarwal,K.
 TITLE Structural analysis of the gene encoding human gastrin: the large
 intron contains an Alu sequence
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 81 (15), 4662-4666 (1984)

THIS PAGE BLANK (USPTO)

Search for 09823101

Your SELECT statement is:

s ssg and cancer

Items File

3 5: Biosis Previews(R)_1969-2003/Mar W2
4 34: SciSearch(R) Cited Ref Sci_1990-2003/Mar W2
4 71: ELSEVIER BIOBASE_1994-2003/Mar W2
30 73: EMBASE_1974-2003/Mar W2
1 98: General Sci Abs/Full-Text_1984-2003/Feb
1 144: Pascal_1973-2003/Mar W1
5 149: TGG Health&Wellness DB(SM)_1976-2003/Feb W4
2 155: MEDLINE(R)_1966-2003/Mar W2
1 159: Cancerlit_1975-2002/Oct
1 162: CAB Health_1983-2003/Jan
2 266: FEDRIP_2003/Jan

SYSTEM:OS - DIALOG OneSearch

File 73:EMBASE 1974-2003/Mar W2

(c) 2003 Elsevier Science B.V.

*File 73: Alert feature enhanced for multiple files, duplicates removal, customized scheduling. See HELP ALERT.

File 159:Cancerlit 1975-2002/Oct

(c) format only 2002 Dialog Corporation

*File 159: Cancerlit ceases updating with immediate effect.

Please see HELP NEWS.

File 155:MEDLINE(R) 1966-2003/Mar W2

(c) format only 2003 The Dialog Corp.

Set Items Description

S1 388 SSG

S2 395880 STOMACH OR GASTRIC

S3 1 S1 AND S2

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 17, 2002, 08:48:53 ; Search time 2206.96 Seconds
(without alignments)
784.882 Million cell updates/sec

Title: US-09-823-101-7
Perfect score: 105
Sequence: 1 tgcagtggtgcacagatcc.....ggccgcctctctgaagctt 105

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapept 1.0

Searched: 1472140 seqs, 8248589755 residues
Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Maximum Match 0%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_da:*
2: gb_bt:*
3: gb_in:*
4: gb_ov:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pt:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vt:*
15: em_da:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_om:*
20: em_or:*
21: em_ov:*
22: em_pat:*
23: em_ph:*
24: em_pl:*
25: em_ro:*
26: em_sts:*
27: em_sy:*
28: em_un:*
29: em_vt:*
30: em_htgo_hum:*
31: em_htgo_inv:*
32: em_htgo_rod:*
33: em_htg_hum:*
34: em_htg_inv:*
35: em_htg_rod:*
36: em_htg_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	47.6	45.3	405	6	E01795	E01795 mRNA sequen
2	47.6	41.3	9		HSRST	V00511 Human mRNA
3	47.6	45.3	829	6	E01796	E01796 DNA sequenc
4	47.6	45.3	829	6	HSRST	X00183 Human gastr
5	47.6	45.3	1217	9	HUMGAST2	X01254 Human gastr
6	47.6	45.3	7739	9	HUMGAST2	M15958 Human gastr
7	47.6	39.6	451	4	ECPPRGAST	Y09440 Equus cabal
8	41.6	36.6	457	4	ECPPRGAST	X15682 Feline mRNA
9	33.6	32.0	447	4	BTTPGAST	M16581 Bovine gastr
10	33.6	32.0	1066	4	BOVGAS	M31657 Bovine gastr
11	32	30.5	462	4	SCGAST	V01303 Porcine mes
12	32	30.5	746	4	OU92801	U92801 Ovis arles
13	30.2	28.8	632	9	H0M2B55H12	AF086156 Homo sapi
14	30.2	28.8	685	9	HSR800786	AL110142 Homo sapi
15	30.2	28.8	1751	9	BC004912	BC004912 Homo sapi
16	30.2	28.8	2857	9	AK023487	AK023487 Homo sapi
17	30.2	28.8	3964	9	AB018271	AB018271 Homo sapi
18	30.2	28.8	67033	9	AL137008	AL137008 Human DNA
19	30	28.6	152366	9	AL138894	AL138894 Human DNA
20	29.8	28.4	32719	2	AC010733	AC010733 Homo sapi
21	29.8	28.4	130000	9	AC087259	AC087259 Homo sapi
22	29.8	28.4	155818	9	AC073504	AC073504 Homo sapi
23	29.8	28.4	167405	8	AP002483	AP002483 Oryza sat
24	29.8	28.4	175565	8	AP003311	AP003311 Oryza sat
25	29.4	28.0	99162	9	AL391361	AL391361 Human DNA
26	29.4	28.0	161079	2	AC083842	AC083842 Homo sapi
27	29.4	28.0	173709	2	AC073167	AC073167 Homo sapi
28	29.4	28.0	182105	2	AC036149	AC036149 Homo sapi
29	29.4	28.0	187098	2	AC015984	AC015984 Homo sapi
30	29.2	27.8	17250	2	AC027264	AC027264 Homo sapi
31	29.2	27.8	190748	9	AC010969	AC010969 Homo sapi
32	29	27.6	208558	2	AL392043	AL392043 Homo sapi
33	28.8	27.4	151	11	G65603	G65603 sy2795 YAC
34	28.8	27.4	153	11	G65864	G65864 sy698 MISCe
35	28.8	27.4	460	10	MMGASTGNI	X94758 M.musculus
36	28.8	27.4	1055	10	MMGASTGNI	X94760 M.musculus
37	28.8	27.4	1267	10	MMU58136	U58136 Mus musculu
38	28.8	27.4	5154	10	MMU34293	U34293 Mus musculu
39	28.8	27.4	137682	9	AC023274	AC023274 Homo sapi
40	28.8	27.4	172811	2	AC080086	AC080086 Homo sapi
41	28.8	27.4	177421	2	AC032013	AC032013 Mus muscu
42	28.8	27.4	177769	9	AC006328	AC006328 Homo sapi
43	28.8	27.4	272351	2	AL590968	AL590968 Mus muscu
44	28.6	27.2	13521	1	AE003904	AE003904 Xylella f
45	28.6	27.2	35240	2	AC040987	AC040987 Homo sapi

ALIGNMENTS

RESULT 1	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	TITLE	AUTHORS	JOURNAL	COMMENT
E01795	E01795	mRNA sequence coding for human gastrin.	E01795	E01795.1	GI:5708448	JP 1989002579-A/1.	Homo sapiens	1	(bases 1 to 405)	Sato, T., Hime, S., Hayashizaki, Y., Katou, K. and Matsubara, K.	RECOMBINANT DNA, MICROORGANISM CONTAINING SAID DNA AND PRODUCTION OF POLYPEPTIDE THEREWITH	Patent: JP 1989002579-A 1 06-JAN-1989;
							Homo sapiens				ATINOMOTO CO INC	OS Homo sapiens (Human)
							Homo sapiens				JP 1989002579-A/1	PD 06-JAN-1989

```

PE 24-JUN-1987 JP 1987156745
PI SATO TAKAKI, HIMENO SEIICHI, HAYASHIZAKI YOSHITIDE, PI
KATOU KIKUYA,
PI MATSUBARA KENICHI
PC C12N15/00.C07H21/04.C12N1/16.C12P21/02.(C12P21/02.C12R1:865);
CC strandedness: Single;
CC topology: Linear;
CC hypothetical: No;
CC anti-sense: No;
CC *source: tissue=Stomach;
CC *source: cell-type=Antrum mucosal tissue;
CC *source: clone=PHG 53;
FH key Location/Qualifiers
FH 5'UTR 1..4
FH sig_peptide 5..61
FH mat_peptide 62..172 /product='human gastrin signal peptide' FT
FT misc_signal 173..178 /product='human gastrin precursor' FT
FT /note='region deleted by tripsin-like FT
FT protease' /note='region deleted by tripsin-like FT
FT mat_peptide 179..307 /product='human gastrin'
FT misc_signal 224..229 /note='region deleted by tripsin-like FT
FT protease' /note='region deleted by tripsin-like FT
FT misc_signal 284..289 /note='region deleted by tripsin-like FT
FT 3'UTR 308..404
FT polyA_signal 378..381
FT polyA_signal 385..390.
location/Qualifiers
1..405
/organism='Homo sapiens'
/db_xref='taxon:9606'
BASE COUNT 99 a 123 c 110 g 73 t
ORIGIN

Query Match 45.3%; Score 47.6; DB 6; Length 405;
Best Local Similarity 92.4%; Pred. No. 1.4e-05;
Matches 61; Conservative 0; Mismatches 4; Indels 1; Gaps 1;

QY 40 gatcgacgacgagtgtgtatgtctgtcctcttcgactagatcagccgctctctg 99
|||||
Db 4 GATCGACGACGACTATGTGTATGTCTGTCTTGCACCT-GGCTCTGCGCGCTTCTCTG 62
|||||

QY 100 aagctt 105
|||||
Db 63 AAGCTT 68

RESULT 2
HSGAST 413 bp mRNA PRI 21-NOV-1994
LOCUS HSGAST 413 bp mRNA PRI 21-NOV-1994
DEFINITION Human mRNA encoding pregastrin (a regulatory hormone of gastric
acid secretion and growth of the gastrointestinal mucosa).
ACCESSION V00511
VERSION V00511.1 GI:31654
KEYWORDS complementary DNA; gastrin; signal peptide.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 413)
Boel,E., Vuust,J., Norris,F., Norris,K., Wind,A., Rehfeld,J.F. and
Matcker,K.A.
Molecular cloning of human gastrin cDNA: evidence for evolution of
gastrin by gene duplication
Proc. Natl. Acad. Sci. U.S.A. 80 (10), 2866-2869 (1983)
JOURNAL 83221503

```

```

COMMENT Data kindly reviewed (21-JUN-1983) by Esper Boel.
FEATURES
source Location/Qualifiers
1..413
/organism='Homo sapiens'
/db_xref='taxon:9606'
mRNA
1..413
/product='pregastrin'
CDS
9..314
/codon_start=1
/product='pregastrin'
/protein_id='CAA23769.1'
/db_xref='GI:31653'
/translaton='MORLCVYVLIIFALALAAFSASWKPDRSQDPAPLGTGANRLLEL
PWLEQGPASHHRRQLGPGPPHLVADPSKKQGPWLEEEEAAGWMDFFRRSAEDEN'
sig_peptide 9..65
polyA_site 413
BASE COUNT 102 a 127 c 110 g 74 t
ORIGIN

Query Match 45.3%; Score 47.6; DB 9; Length 413;
Best Local Similarity 92.4%; Pred. No. 1.4e-05;
Matches 61; Conservative 0; Mismatches 4; Indels 1; Gaps 1;

QY 40 gatcgacgacgagtgtgtatgtctgtcctcttcgactagatcagccgctctctg 99
|||||
Db 8 GATCGACGACGACTATGTGTATGTCTGTCTTGCACCT-GGCTCTGCGCGCTTCTCTG 66
|||||

QY 100 aagctt 105
|||||
Db 67 AAGCTT 72

RESULT 3
LOCUS E01796 829 bp DNA PAT 29-SEP-1997
DEFINITION DNA sequence coding for human gastrin.
ACCESSION E01796
VERSION E01796.1 GI:2170048
KEYWORDS JP 1989002579-A/2.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 829)
Sato,T., Himeno,S., Hayashizaki,Y., Katou,K. and Matsubara,K.
RECOMBINANT DNA, MICROORGANISM CONTAINING SAID DNA AND PRODUCTION
OF POLYPEPTIDE THEREWITH
Patent: JP 1989002579-A 2 06-JAN-1989.
JOURNAL AJINOMOTO CO INC
OS Homo sapiens (Human)
PN JP 1989002579-A/2
PD 06-JAN-1989
PI 24-JUN-1987 JP 1987156745
PI SATO TAKAKI, HIMENO SEIICHI, HAYASHIZAKI YOSHITIDE, PI
KATOU KIKUYA,
PI MATSUBARA KENICHI
PC C12N15/00.C07H21/04.C12N1/16.C12P21/02.(C12P21/02.C12R1:865);
CC strandedness: Double;
CC topology: Linear;
CC hypothetical: No;
CC anti-sense: No;
CC *source: tissue=Stomach;
CC *source: cell-type=Antrum mucosal tissue;
CC *source: clone=plamdaHG 11;
FH key Location/Qualifiers
FH 5'UTR 1..263
FH sig_peptide 34..42
FH misc_signal 64..69
FT /standard_name='GCCAATCT model sequence' FT
FT misc_signal 64..69 /standard_name='Gold berg-Hogness box' FT

```



```

CDS          join(264..474,605..699)
            /product='human gastrin'
FT           sig_peptide 264..437
FT           mat_peptide join(438..474,605..699)
FT           3'UTR       700..831
FEATURES
  source      Location/Qualifiers
            1..829
            /organism='Homo sapiens'
            /db_xref='taxon:9606'
BASE COUNT   199 a 250 c 213 g 167 t
ORIGIN
Query Match   45.3% Score 47.6; DB 6; Length 829;
Best Local Similarity 92.4%; Pred. No. 1.2e-05;
Matches 61; Conservative 0; Mismatches 4; Indels 1; Gaps 1;

Qy 40 gatgcagcagatgtgtatgtctgtctcttccttcagatcagcagcgcctctctg 99
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 261 GATGAGCGACTATGTGTATGTCGATCTTTCACCT-GGCTCTGGCCGCTTCTCTG 319

Qy 100 aagctt 105
    ||||||
Db 320 AAGCTT 325

RESULT 4
HSGAS1 LOCUS HSGAS1 829 bp DNA PRI 10-FEB-1999
DEFINITION Human gastrin gene.
ACCESSION X00183
VERSION X00183.1 GI:31648
KEYWORDS gastrin.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 829)
AUTHORS Kato, K., Hayashizaki, Y., Takahashi, Y., Himenos, S. and Matsubara, K.
TITLE Molecular cloning of the human gastrin gene
JOURNAL Nucleic Acids Res. 11 (23), 8197-8203 (1983)
MEDLINE 84169471
FEATURES
  source      Location/Qualifiers
            1..829
            /organism='Homo sapiens'
            /db_xref='taxon:9606'
            34..42
            /note='variant of CAAT-box'
            64..69
            /note='Goldberg-Hogness box'
            151..472
            /note='exon 1 (472 is 1st base in codon)'
            151..796
            /note='putative primary transcript'
            join(262..472,603..697)
            /note='precursor'
            /codon_start=1
            /protein_id='CAA25005.1'
            /db_xref='GI:31649'
            /db_xref='SWISS-PROT:P01350'
            /translation='MORLCYVILFALALAFSAKPPSOQDPAFLGTGANRDEL
            PWLEQGPASHHRRLCPQGPPLVADPSKKQGPWLEEEBAYGMDFGRSAEDEN'
            262..318
            join(436..472,603..667)
            /note='protein sequence is in conflict with the conceptual
            translation'
            /codon_start=1
            /product='G34 (big gastrin)'
            /protein_id='CAA25006.1'
            /db_xref='GI:1335072'
            /db_xref='SWISS-PROT:P01350'
sig_peptide
CDS

```

```

intron
      /translation='QLGPGPPLVADPSKKQGPWLEEEBAYGMDF'
      473..602
      /note='intron'
      603..796
      /note='exon 2 (603 is 2nd base in codon)'
      617..667
      /note='protein sequence is in conflict with the conceptual
      translation'
      /codon_start=1
      /product='G17 (little gastrin) with full biological
      activity'
      /protein_id='CAA25007.1'
      /db_xref='GI:1335073'
      /db_xref='SWISS-PROT:P01350'
      /translation='QGPWLEEEBAYGMDF'
      793
      /note='putative polyadenylation site'
      797
      /note='putative polyadenylation site'
      797
      /note='putative polyadenylation site'
BASE COUNT   198 a 251 c 213 g 167 t
ORIGIN
Query Match   45.3% Score 47.6; DB 9; Length 829;
Best Local Similarity 92.4%; Pred. No. 1.2e-05;
Matches 61; Conservative 0; Mismatches 4; Indels 1; Gaps 1;

Qy 40 gatgcagcagatgtgtatgtctgtctcttccttcagatcagcagcgcctctctg 99
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 261 GATGAGCGACTATGTGTATGTCGATCTTTCACCT-GGCTCTGGCCGCTTCTCTG 319

Qy 100 aagctt 105
    ||||||
Db 320 AAGCTT 325

RESULT 5
HMGAST2 LOCUS HMGAST2 1217 bp DNA PRI 08-AUG-1995
DEFINITION Human gastrin gene, complete cds.
ACCESSION K01254 J00147
VERSION K01254.1 GI:182987
KEYWORDS Alu repeat; gastrin; repeat region.
SEGMENT 2 of 2
SOURCE Human fetal liver genomic DNA (Charon 4A library of Lawn et al.),
and gastrinoma, cDNA to mRNA (see comment).
Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 601 to 815; 946 to 1136)
AUTHORS Kato, K., Himeno, S., Takahashi, Y., Wakabayashi, T., Tarui, S. and
Matsubara, K.
TITLE Molecular cloning of human gastrin precursor cDNA
JOURNAL Gene 26 (1), 53-57 (1983)
MEDLINE 84159488
REFERENCE 2 (bases 600 to 815; 946 to 1139)
AUTHORS Boel, E., Vunst, J., Norris, F., Norris, K., Wind, A., Rehfeld, J.F. and
Marcker, K.A.
TITLE Molecular cloning of human gastrin cDNA: evidence for evolution of
gastrin by gene duplication
Proc. Natl. Acad. Sci. U.S.A. 80 (10), 2866-2869 (1983)
JOURNAL 83221503
REFERENCE 3 (bases 471 to 1151)
AUTHORS Wborg, O., Berglund, L., Boel, E., Norris, F., Norris, K.,
Rehfeld, J.F., Marcker, K.A. and Vunst, J.
TITLE Structure of a human gastrin gene
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 81 (4), 1067-1069 (1984)
MEDLINE 84144842
REFERENCE 4 (bases 1 to 1217)
AUTHORS Ito, R., Sato, K., Helmer, T., Jay, G. and Agarwal, K.
TITLE Structural analysis of the gene encoding human gastrin: the large
intron contains an Alu sequence
Proc. Natl. Acad. Sci. U.S.A. 81 (15), 4662-4666 (1984)
JOURNAL

```

MEDLINE 84272693

COMMENT The gastrin-34 protein is also known as big gastrin and the gastrin-17 form as little gastrin. Other portions of the propeptide may have hormonal or enzymatic activity. [2] discusses the internal homology between bases 699-776 and 798-875 below in terms of gene duplication. Both [2] and [1] compare human and porcine sequences.

Complete source information:
Human fetal liver genomic DNA (Charon 4A library of Lawn et al.)
[1],[3],[4] and gastrinoma, cDNA to mRNA [2],[1].
Location/Qualifiers
1. 1217

FEATURES source

/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="17q"
join(K01253.1:114..395,1..1040)
/gene="GAS"
prim_transcript <1..1138
/note="GAS"
prim_transcript <1..1136
/note="gastrin mRNA (alt.) [3]"
<1..599
/gene="GAS"
intron
/note="gastrin mRNA (alt.) [1],[3]"
prim_transcript <1..1139
/note="gastrin mRNA intron A"
repeat_region
171..499
/note="Alu repeat"
601
/gene="GAS"
/citation=[1]
/replace=""
605..661
/gene="GAS"
/note="preprogastrin signal peptide"
605..815
/gene="GAS"
exon
/note="preprogastrin (first expressed); G00-119-261"
<605..815
/gene="GAS"
exon
/note="first expressed; G00-119-261"
join(605..815,946..1040)
/gene="GAS"
/note="preprogastrin"
/codon_start=1
/protein_id="AAB59533.1"
/db_xref="GI:182989"
/translation="MQRICVYVILFALALAAFSASMKPRSQPDAPLIGANRDLTL
PMLEQGPASHHRQLGPGCPHLVADPSKKQCPMLEEEEAATGMNDFGRSAEDER"
779..1010
/gene="GAS"
mat_peptide
/note="gastrin g-34"
816..945
/gene="GAS"
intron
/note="gastrin intron B"
861..863
/gene="GAS"
conflict
/citation=[3]
/replace=""
946..>1040
/gene="GAS"
exon
/note="second expressed; G00-119-261"
946..1040
/gene="GAS"
mat_peptide
/note="preprogastrin (second expressed)"
960..1010
/gene="GAS"
conflict
/note="gastrin g-17"
1087
/citation=[3]
/replace=""
1108..1109
/citation=[1]
/replace=""

conflict

1131
/citation=[1]
/replace=""

BASE COUNT 309 a 332 c 336 g 240 t
ORIGIN About 3.2 kb after segment 1.

Query Match 45.3%; Score 47.6; DB 9; Length 1217;
Best Local Similarity 92.4%; Pred. No. 1.2e-05;
Matches 61; Conservative 0; Mismatches 4; Indels 1; Gaps 1;

OY 40 gatgcagcagcagatggtatgtatgtcgtcttgcactgagcagccgctctctg 99
|||||
Db 604 GATCAGCAGCTATGTGTATGTGATGTGACT-GGCTCTGCCGCTCTCTG 662
OY 100 aagctt 105
|||||
Db 663 AAGCTT 668

RESULT 6
LOCUS HUMGASTA 7739 bp DNA PRI 08-NOV-1994
DEFINITION Human gastrin gene, complete cds.
ACCESSION M15958
VERSION M15958.1 GI:182990
KEYWORDS Alu repeat; gastrin; repeat region.
SOURCE Human gastric antrum and gastrinoma DNA.
ORGANISM Homo sapiens

REFERENCE
AUTHORS Kariya,Y., Kato,K., Hayashizaki,Y., Himeno,S., Tarui,S. and Matsubara,K.
TITLE Expression of human gastrin gene in normal and gastrinoma tissues
JOURNAL Gene 50 (1-3), 345-352 (1986)
MEDLINE 87219893
COMMENT Draft entry and printed copy of sequence in [1] kindly provided by Y.Kariya, 08-APR-1987. The IVS B seems to end with bases ac and not with ag, as would be expected. Letter sent to the author to clarify, 14-MAR-1988. Letter received from the author, 06-APR-1988. Base at position 6629 changed from c to g.

FEATURES

source
1..7739
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="17q"
repeat_region
377..707
/note="Alu-1 repeat"
991..1326
/note="Alu-2 repeat"
2074..2390
/note="Alu-3 repeat"
prim_transcript 3183..6821
/note="gastr mRNA (alt.)"
prim_transcript 3186..6821
/note="gastr mRNA (alt.)"
prim_transcript 3188..6821
/note="gastr mRNA (alt.)"
intron
3243..6283
/note="gastr mRNA intron A"
repeat_region
3775..4071
/note="Alu-4 repeat"
4251..4585
/note="Alu-5 repeat"
5014..5328
/note="Alu-6 repeat"
5343..5643
/note="Alu-7 repeat"
5861..6184
/note="Alu-8 repeat"
exon
<6289..6499
/gene="GAS"

gene /note="gastrin, (first expressed exon); G00-119-261"
/number=2
6289. .6499
/gene="GAS"
CDS join(6289. .6499,6630. .6724)
/note="gastrin"
/codon_start=1
/protein_id="AA52520.1"
/db_xref="GI:182991"
Intron /translation="MORLCYVILFALALAFSEASMKPPSOQPDAPLGTCANDLEL
PWLEQGPASHHRQLDGPQPPHIVADPSKQGPWLEEEBAAYGMDPGRSAEDEN
6500. .6629
/note="gastrin cds intron B"
6630. .>6724
/note="gastrin"
exon /number=3
7005. .7349
repeat_region /note="Alu-9 repeat"
BASE COUNT 1964 a 2036 c 1844 g 1895 t
ORIGIN 1 bp upstream of EcoRI site; chromosome 17q.

Query Match 45.3%; Score 47.6; DB 9; Length 7739; *
Best Local Similarity 92.4%; Pred. No. 9.5e-06;
Matches 61; Conservative 0; Mismatches 4; Indels 1; Gaps 1;

OY 40 gatcagcagcagatgtgtatgtctgtctcttgcactagatcagccgctctctg 99
|||||
Db 6288 GATCAGCGACTATGTGTATGTCTGACTCTTGCCTGGCGCTTCTCTG 6346
|||||

OY 100 aagctt 105
|||||
Db 6347 AAGCTT 6352

RESULT 7
ECPPGAST
LOCUS ECPPPGAST 451 bp mRNA MAM 19-AUG-1998
DEFINITION Equus caballus mRNA for preprogastrin.
ACCESSION Y09440
VERSION Y09440.1 GI:3445273
KEYWORDS gast gene; preprogastrin.
SOURCE horse.
ORGANISM Equus caballus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
REFERENCE 1 (bases 1 to 451)
Johnsen, A.H., Sandin, A., Rourke, I.J., Bundgaard, J.R., Nilsson, G.
and Rehfeld, J.F.
Unique progastrin processing in equine G-cells suggests marginal
tyrosyl sulfotransferase activity
Eur. J. Biochem. 255 (2), 432-438 (1998)

TITLE
JOURNAL MEDLINE
REFERENCE 2 (bases 1 to 451)
Rourke, I.J.
AUTHORS Rourke, I.J.
TITLE Direct Submission
JOURNAL Submitted (14-NOV-1996) I. J. Rourke, Department of Clinical
Biochemistry, National University Hospital, Blegdamsvej 9,
Copenhagen, DK-2100, DENMARK
REMARK 3 (bases 1 to 451)
Revised by [3]
REFERENCE 3 (bases 1 to 451)
Rourke, I.J.
TITLE Direct Submission
JOURNAL Submitted (17-AUG-1998) I. J. Rourke, Department of Clinical
Biochemistry, National University Hospital, Blegdamsvej 9,
Copenhagen, DK-2100, DENMARK
COMMENT On Aug 20, 1998 this sequence version replaced gi:1684682.
FEATURES
source location/Qualifiers
1. .451
/organism="Equus caballus"
/db_xref="taxon:9796"
/dev_stage="adult"
/tissue_type="antral mucosa"

gene 41. .364
/gene="gast"
41. .364
CDS /gene="gast"
/codon_start=1
/product="preprogastrin"
/protein_id="CAA70590.1"
/db_xref="GI:1684683"
mat_peptide /db_xref="SWISS-PROT:P55885"
/translation="MRRLCYVILFALALAFSEASMKPPSOQPDAPLGTCANDLEL
HWPQDLRLGPASHHRQLDGPQPPHIVADPSKQGPWLEEEBAAYGMDPGRSAEDEN
GDQSP"
224. .328
/evidence="experimental"
/product="mature gastrin-34 coding sequence including
amidation substrate"
429. .434
polya_signal
BASE COUNT 102 a 142 c 127 g 80 t
ORIGIN

Query Match 39.6%; Score 41.6; DB 4; Length 451;
Best Local Similarity 85.3%; Pred. No. 0.0015;
Matches 58; Conservative 0; Mismatches 9; Indels 1; Gaps 1;

OY 38 atgacgacgacgacgtgtatgtctgtctcttgcactagatcagccgctctc 97
|||||
Db 38 AAGATGCGCGACTGTGTATGTCTGATCTTGGCTGGCGGCGCTTCTC 96
|||||

OY 98 tgaagctt 105
|||||
Db 97 CGAAGCTT 104

RESULT 8
FSPPGAST
LOCUS FSPPGAST 457 bp mRNA MAM 19-MAR-1991
DEFINITION Feline mRNA for preprogastrin.
ACCESSION X16582
VERSION X16582.1 GI:1099
KEYWORDS gastrin; preprogastrin.
SOURCE wild cat.
ORGANISM Felis silvestris
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Carnivora; Fissipedia; Felidae; Felis.
REFERENCE 1 (bases 1 to 457)
Kim, S.J.
AUTHORS Kim, S.J.
TITLE Direct Submission
JOURNAL Submitted (22-SEP-1989) Kim S. J., Korea Advanced Institute of
Science and Technology, Department of Biological Science and
Engineering, P.O. Box 150, Chongryang, Seoul 131, Korea
REFERENCE 2 (bases 1 to 457)
Kim, S.J., Uhm, K.N., Kang, Y.K. and Yoo, O.J.
Bovine and feline gastrin cDNA sequences and the amino acid and
nucleotide sequence homologies among mammalian species
JOURNAL DNA Seq. 1 (3), 181-187 (1991)
MEDLINE 92127058
FEATURES
source location/Qualifiers
1. .457
/organism="Felis silvestris"
/db_xref="taxon:9683"
/dev_stage="adult"
47. .103
47. .361
/codon_start=1
/product="preprogastrin"
/protein_id="CAA34599.1"
/db_xref="GI:1100"
/db_xref="SWISS-PROT:P01354"
/translation="MORLCYVILFALALAFSEASMKPPSOQPDAPLGTCANDLEL
HWPQDLRLGPASHHRQLDGPQPPHIVADPSKQGPWLEEEBAAYGMDPGRSAEDEN
HP"

Db 596 CGAAGCTT 603

RESULT 11

SCGAST 462 bp mRNA MAM 01-OCT-1996
 LOCUS Porcine messenger RNA for preprogastrin.
 DEFINITION V01303.1 J00651
 ACCESSION V01303.1 GI:1846
 VERSION Complementary DNA; gastrin.
 KEYWORDS
 SOURCE
 ORGANISM Sus scrofa

REFERENCE

1 (bases 1 to 462)
 Yoo, O.-J., Powell, C.T. and Agarwal, K.L.
 Molecular cloning and nucleotide sequence of full-length of cDNA coding for porcine gastrin
 Proc. Natl. Acad. Sci. U.S.A. 79 (4), 1049-1053 (1982)

FEATURES

Location/Qualifiers
 1..462
 /organism="Sus scrofa"
 /db_xref="taxon:9823"
 1..462
 /note="messenger RNA"
 62..376
 /note="reading frame preprogastrin"
 /codon_start=1
 /protein_id="CAA24610.1"
 /db_xref="GI:1847"
 /translation="MQRICAVLIHLVLAALACASASWKPGFOLDASSGPGANGKEP
 HELDRLGPASHRRRLGIQDPPLVADLAKKQGWMEEEEAAYGMDFGRSAEEGDQ
 RP"

BASE COUNT 110 a 147 c 135 g 70 t
 ORIGIN

Query Match 30.5%; Score 32; DB 4; Length 462;
 Best Local Similarity 76.5%; Pred. No. 2.9;
 Matches 52; Conservative 0; Mismatches 15; Indels 1; Gaps 1;
 Oy 38 atgatgcagcagcgtatgtgtatgtctgtccttgcactagatcagccgcctctc 97
 Db 59 AAGATGCAGCGCTGCTGCGCTATGCTCGATCGATGCTGCTGCTGCTGCTG 117
 Oy 98 tgaagctt 105
 Db 118 TGAAGCTT 125

RESULT 12

CAU92801 746 bp DNA MAM 16-JUL-1999
 LOCUS Ovis aries preprogastrin (GAS) gene, exons 2, 3 and complete cds.
 DEFINITION U92801
 ACCESSION U92801.1 GI:1917024
 VERSION
 KEYWORDS
 SOURCE sheep.
 ORGANISM Ovis aries

REFERENCE

1 (bases 1 to 746)
 Moore, C., Jie, R., Shulkes, A. and Baldwin, G.S.
 Molecular cloning and sequence of the ovine gastrin gene
 JOURNAL DNA Seq. 8 (1-2), 39-44 (1997)
 MEDLINE 9522119
 PUBMED 2 (bases 1 to 746)
 REFERENCE Moore, C., Ren, J., Shulkes, A. and Baldwin, G.S.
 AUTHORS

TITLE Direct Submission
 JOURNAL Submitted (10-MAR-1997) Surgery, ASRMC, Studley Road, Heidelberg,
 Vic. 3084, Australia
 FEATURES Location/Qualifiers
 source 1..746
 /organism="Ovis aries"
 /db_xref="taxon:9940"
 /cell_type="leukocytes"
 <1..>746
 /gene="GAS"
 <1..99
 /gene="GAS"
 /number=1
 /number=1
 /join(<100..315,462..>647)
 /gene="GAS"
 /product="preprogastrin"
 100..315
 /gene="GAS"
 /number=2
 /join(105..315,462..565)
 /gene="GAS"
 /note="precursor of the hormone gastrin"
 /codon_start=1
 /product="preprogastrin"
 /protein_id="AAB51307.1"
 /db_xref="GI:1917025"
 /translation="MQRICAVLIHLVLAALACASASWKPGFOLDASSGPGANGKEP
 LRMDRLGPASHRRRLGIQDPPLVADLAKKQGWMEEEEAAYGMDFGRSAEEGDQ
 HP"
 316..461
 /gene="GAS"
 /number=2
 462..>647
 /gene="GAS"
 /number=3

BASE COUNT 150 a 241 c 221 g 134 t
 ORIGIN

Query Match 30.5%; Score 32; DB 4; Length 746;
 Best Local Similarity 76.5%; Pred. No. 2.7;
 Matches 52; Conservative 0; Mismatches 15; Indels 1; Gaps 1;
 Oy 38 atgatgcagcagcgtatgtgtatgtctgtccttgcactagatcagccgcctctc 97
 Db 102 AAGATGCAGCGCTGCTGCGCTATGCTCGATCGATGCTGCTGCTGCTG 160
 Oy 98 tgaagctt 105
 Db 161 CGAAGCTT 168

RESULT 13

HUMZB55H12 632 bp mRNA PRI 29-AUG-1998
 LOCUS Homo sapiens full length insert cDNA clone ZB55H12.
 DEFINITION AF086156
 ACCESSION AF086156.1 GI:3483501
 VERSION
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE

1 (bases 1 to 632)
 Moessner, J., Tan, F., Marra, M., Kucaba, T., Yandell, M., Martin, J.,
 Marth, G., Bowles, L., Wylie, T., Bowers, Y., Steptoe, M., Theising, B.,
 Geisel, S., Allen, M., Underwood, K., Chappell, J., Person, B.,
 Gibbons, M., Harvey, N., Pape, D., Chamberlain, A., Morales, R.,
 Schurk, R., Rilter, E., Kohn, S., Swaller, T., Behmer, K., Hillier, L.,
 Wilson, R. and Waterson, R.
 Full Clone Sequencing of the longest Available Member from Each
 Unigene Cluster
 JOURNAL unpublished

REFERENCE 2 (bases 1 to 632)
 AUTHORS Waterston,R.
 TITLE Direct Submission
 JOURNAL Submitted (24-AUG-1998) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
 COMMENT SUBMITTED BY: Genome Sequencing Center
 Department of Genetics
 Washington University
 St. Louis MO 63108, USA
 http://genome.wustl.edu/gsc
 mailto:est@watson.wustl.edu

NOTICE: This sequence represents the full insert of this cDNA. No attempt has been made to verify whether this corresponds to the full-length of the original mRNA from which it was derived. We have tried to obtain double-stranded, or double chemistry sequence across the entire clone, but potentially, there are areas in the sequence where this level of coverage was not achieved. Nevertheless, we are confident of the accuracy of this sequence as all regions of low quality, as defined by PHRAP (P. Green, in preparation), were visually inspected and edited accordingly. The consensus quality values for this sequence have been submitted separately.

The location of this clone is unknown.

FEATURES
 source Location/Qualifiers
 1..632
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="Soares_fetal_lung_NbHL19W"
 /clone="IMAGE:307559"

BASE COUNT 183 a 117 c 123 g 209 t
 ORIGIN

Query Match 28.8%; Score 30.2; DB 9; Length 632;
 Best Local Similarity 69.5%; Pred. No. 11;
 Matches 41; Conservative 0; Mismatches 18; Indels 0; Gaps 0;
 Oy 15 acatccgcagctctctgtacagatgacgagcagatgtgtatgtctgtcttt 73
 || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 281 ACTTCCTGTGAGTCTCTGTGACACTGCAAGACATGCGTGGTGTGTTGTTCTTT 339

RESULT 14
 LOCUS HSM800786 685 bp mRNA PRI 18-FEB-2000
 DEFINITION Homo sapiens mRNA: cDNA DKFZp564D0264 (from clone DKFZp564D0264).
 ACCESSION AL110142
 VERSION AL110142.1 GI:5817037
 KEYWORDS

SOURCE

human.
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 685)
 Duesterhoeft,A., Lauber,J., Mewes,H.W., Gassenhuber,J. and Wiemann,S.

TITLE Direct Submission
 JOURNAL Submitted (15-AUG-1999) MIPS, Am Klopferstr. 18a, D-82152 Martinsried, GERMANY
 COMMENT Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email: s.wiemann@dkfz-heidelberg.de; sequenced by Olagen (Hilden/Germany) within the cDNA sequencing consortium of the German Genome Project.

This clone (DKFZp564D0264) is available at the RZPD in Berlin. Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clonegrzpd.de Further information about the clone and the sequencing project is available at <http://www.mips.biochem.mpg.de/proj/cDNA/>.

FEATURES
 source Location/Qualifiers
 1..685

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="DKFZp564D0264"
 /clone_lib="564 (synonym: hfbz2). Vector pAMP1; host X1-2blue; sites NotI + SalI"
 /dev_stage="fetal"
 /tissue_type="brain"
 polyA_signal 648..653
 polyA_site 666
 BASE COUNT 239 a 107 c 112 g 227 t
 ORIGIN

Query Match 28.8%; Score 30.2; DB 9; Length 685;
 Best Local Similarity 69.5%; Pred. No. 11;
 Matches 41; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

Oy 15 acatccgcagctctctgtacagatgacgagcagatgtgtatgtctgtcttt 73
 || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 170 ACTTCCTGTGAGTCTCTGTGACACTGCAAGACATGCGTGGTGTGTTGTTCTTT 112

RESULT 15
 LOCUS BC004912/c 1751 bp mRNA PRI 12-JUL-2001
 DEFINITION Homo sapiens, clone IMAGE:3534745, mRNA, partial cds.
 ACCESSION BC004912
 VERSION BC004912.1 GI:13436208
 KEYWORDS

SOURCE

human.
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 1751)
 Strausberg,R.

TITLE Direct Submission
 JOURNAL Submitted (21-MAR-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
 REMARK
 COMMENT NIH-MGC Project URL: <http://mgc.ncl.nih.gov>
 Contact: MGC help desk
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: DCTD/DRP
 CDNA Library Preparation: Rubin Laboratory
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Institute for Systems Biology
<http://www.systemsbio.org>
 contact: amadan@systemsbiology.org
 Anup Madan, Rachel Dickhoff, Jessica Fahey, Stephanie Ford, Julia Greene, Mark Kettelman and Anuradha Madan

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>
 Series: IRAL Plate: 11 Row: a Column: 4
 This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis.

FEATURES
 source Location/Qualifiers
 1..1751
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:3534745"
 /tissue_type="lung, small cell carcinoma"
 /clone_lib="NIH_MGC_7"
 /lab_host="DH10B-R"
 /note="Vector: pOTB7"
 <1..788
 /codon_start=3
 /product="unknown (protein for IMAGE:3534745)"
 /protein_id="AAH04912.1"
 /db_xref="GI:13436209"
 /translation="HEVROVAKCKAKRFQVEIGIDNRYREGDSQQLRLVRLNSTV

CDS
 1..788
 /product="unknown (protein for IMAGE:3534745)"
 /protein_id="AAH04912.1"
 /db_xref="GI:13436209"
 /translation="HEVROVAKCKAKRFQVEIGIDNRYREGDSQQLRLVRLNSTV

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 17, 2002, 06:42:46 : Search time 2644.98 Seconds
(without alignments)
438.772 Million cell updates/sec

Title: US-09-823-101-6
Perfect score: 108
Sequence: 1 acctattgtatattgtgaga.....aaaaaaaaaaaaaacgcgcg 108

Scoring table:
IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues
Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: em_estfun:*
2: em_esthum:*
3: em_estlin:*
4: em_estlom:*
5: em_estlpl:*
6: em_estlba:*
7: em_estlro:*
8: em_estlov:*
9: em_estc:*
10: gb_estl:*
11: gb_estl2:*
12: gb_hlc:*
13: gb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_inv:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_rtd:*
20: em_gss_vrt:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	87	80.6	219	10	AM022375 df37g02.y
2	87	80.6	392	10	AM022874 df45d04.y
3	87	80.6	447	10	AM022874 df45d04.y
4	85.4	79.1	409	11	N68114 y254a12.s1
5	85.4	79.1	591	10	AM069257 cr42c12.x
6	84.6	78.3	452	10	AM022830 df44f07.y
7	84.4	78.1	184	10	AM022830 df44f07.y
8	83.8	77.6	433	10	AM022830 df44f07.y
9	83.8	77.6	543	10	AM022830 df44f07.y
10	83.6	77.4	275	11	BI339163 364014.MA
11	83	76.9	510	10	AI620752 t096f12.x
12	82.4	76.3	520	10	AI435232 t110d11.x

Result No.	Score	Query Match	Length	DB ID	Description
13	82.2	76.1	529	10	AI697004 w676g06.x
14	82	75.9	529	10	AM958385 EST370455
15	82	75.9	547	10	AI120039 AI120039
16	81.8	75.7	291	10	AI753826 cr17b05.x
17	81.4	75.4	408	10	AI754101 cr19g05.x
18	81	75.0	176	10	AI632352 cr22h07.x
19	80.4	74.4	366	10	AM021243 df20g02.y
20	79.8	73.9	347	10	AI433059 ch37c11.x
21	79.4	73.5	474	10	AI610692 tpe4012.x
22	79.2	73.3	588	10	AM087960 xda5d12.x
23	78.8	73.0	448	10	AI961907 w40f10.x
24	78.8	73.0	521	10	AI955544 w58a01.x
25	78.4	72.6	593	10	AI801341 t088a06.x
26	78	72.2	440	10	AM022916 cr29e09.x
27	78	72.2	562	10	AI754743 cr29e09.x
28	77.4	71.7	478	10	AI923220 x184d02.x
29	77	71.3	356	10	AI872665 w08c07.x
30	77	71.3	397	10	AI753154 cr06b02.x
31	77	71.3	407	10	AI754915 cr32f08.x
32	77	71.3	507	10	AM069076 cr41a05.x
33	77	71.3	507	10	AM069409 cr45e05.x
34	76.8	71.1	239	11	BF63258 BF63258
35	76.4	70.7	973	11	BG571244 BG571244
36	76	70.4	143	10	AI553940 t054c01.x
37	76	70.4	388	10	AM069707 cr51h10.x
38	76	70.2	670	11	BG939476 cr51h10.x
39	75.8	70.2	518	10	AI569311 cr78g06.x
40	75.4	69.8	519	10	AI955867 w6f0f09.x
41	75	69.4	379	10	AM069320 cr44b08.x
42	75	69.4	466	10	AI753632 cr13b08.x
43	75	69.4	467	10	AI754595 cr27d07.x
44	75	69.4	495	10	AI074234 cr285b09.x
45	75	69.4	509	10	AI520725 t106g07.x

ALIGNMENTS

RESULT 1
AM022375 219 bp mRNA EST 13-SEP-1999
LOCUS df37g02.y1 Morton Fetal Cochlea Homo sapiens CDNA clone
DEFINITION IMAGE:2485802 5', mRNA sequence.
ACCESSION AM022375
VERSION AM022375.1 GI:5675905
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 219)
Robertson,N.G., Khetarpal,U., Gutierrez-Espeleta,G.A., Bieber,F.R.
and Morton,C.C.
Isolation of novel and known genes from a human fetal cochlear CDNA
library using subtractive hybridization and differential screening
Genomics 23, 42-50 (1994)
95130111

JOURNAL MEDLINE

CONTACT: Morton, C. C.
Departments of Pathology and Obstetrics, Gynecology and
Reproductive Biology
Brigham and Women's Hospital
75 Francis Street, Harvard Medical School, Boston, MA 02115, USA
Tel: 617 732 7980
Fax: 617 738 6996
Email: cmorton@rics.bwh.harvard.edu
DNA sequencing and analyses were performed by National Institutes
of Health Intramural Sequencing Center (NISC); see
http://www.nisc.nih.gov/.
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lln.gov) for further information.
Plate: LLM6185 row: M column: 3
Seq primer: M13RP1 reverse primer (ABI).
Location/Qualifiers

FEATURES

```

source
1. .219
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2485802"
/clone_lib="Morton Retal Cochlea"
/tissue_type="cochlea"
/dev_stage="16-22 week fetus"
/lab_host="SOLR cells (kanamycin resistant)"
/notes="Organ: ear; Vector: pBluescript SK-; Site: 1; EcorI; Site 2: XhoI; Reference: Genomics 23, 42-50 (1994) Cloned unidirectionally. Primer: Oligo dT. Fetal cochlea, normal. 37% of inserts <0.5 kb, 56% 0.5-1.0 kb, 7% >1 kb. Uni-ZAP XR Vector. Library constructed by N. Robertson, C. Morton. -5' adaptor sequence: 5' GAATTCGACGACG 3' -3' adaptor sequence: 5' CTCGAGTCTTTTCTTTTCTTTT 3'."

BASE COUNT      93 a      28 c      32 g      66 t
ORIGIN

Query Match      80.6%; Score 87; DB 10; Length 219;
Best Local Similarity 99.0%; Pred. No. 0.4;
Matches 98; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 6 ttgtatattgagatgtttaataaa-tgtgaaaaaatgaataaagcatgttggct 64
|||||
Db 113 TTGTATATGTGAGATGTTTAAATTAATGTGAAAAAATGAATTAAGCATGTTGGTT 172

QY 65 ttccaaagaagaaaaaagaaaaaagaaaaaagaaaaa 103
|||||
Db 173 TTCCAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAA 211

RESULT 2
LOCUS      AM022874      392 bp      mRNA      EST      13-SEP-1999
DEFINITION d145804.y1 Morton Fetal Cochlea Homo sapiens cDNA clone
VERSION     AM022874
XREF        AM022874.1 GI:5876404
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
REFERENCE   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS     Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 392)
Robertson, N.G., Khatripal, U., Gutierrez-Espeleta, G.A., Bieber, F.R.
and Morton, C.C.
Isolation of novel and known genes from a human fetal cochlear cDNA
library using subtractive hybridization and differential screening
Genomics 23, 42-50 (1994)
95130111
Contact: Morton, C. C.
Departments of Pathology and Obstetrics, Gynecology and
Reproductive Biology
Brigham and Women's Hospital
75 Francis Street, Harvard Medical School, Boston, MA 02115, USA
Tel: 617 732 7980
Fax: 617 738 6996
Email: cmorton@rics.bwh.harvard.edu
DNA sequencing and analyses were performed by National Institutes
of Health Intramural Sequencing Center (NISC; see
http://www.nisc.nih.gov).
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Plate: L1A6187 row: G column: 7
Seq primer: M13RP1 reverse primer (ABI).
Location/Qualifiers
1. .392
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2486430"
/clone_lib="Morton Retal Cochlea"
/tissue_type="cochlea"

```

```

/dev_stage="16-22 week fetus"
/lab_host="SOLR cells (kanamycin resistant)"
/notes="Organ: ear; Vector: pBluescript SK-; Site: 1; EcorI; Site 2: XhoI; Reference: Genomics 23, 42-50 (1994) Cloned unidirectionally. Primer: Oligo dT. Fetal cochlea, normal. 37% of inserts <0.5 kb, 56% 0.5-1.0 kb, 7% >1 kb. Uni-ZAP XR Vector. Library constructed by N. Robertson, C. Morton. -5' adaptor sequence: 5' GAATTCGACGACG 3' -3' adaptor sequence: 5' CTCGAGTCTTTTCTTTTCTTTT 3'."

BASE COUNT      139 a      58 c      66 g      127 t      2 others
ORIGIN

Query Match      80.6%; Score 87; DB 10; Length 392;
Best Local Similarity 99.0%; Pred. No. 0.26;
Matches 98; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 6 ttgtatattgagatgtttaataaa-tgtgaaaaaatgaataaagcatgttggct 64
|||||
Db 261 TTGTATATGTGAGATGTTTAAATTAATGTGAAAAAATGAATTAAGCATGTTGGTT 320

QY 65 ttccaaagaagaaaaaagaaaaaagaaaaaagaaaaa 103
|||||
Db 321 TTCCAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAA 359

RESULT 3
LOCUS      AM955633      447 bp      mRNA      EST      01-JUN-2000
DEFINITION EST367703 MAGE resequences, MAGD Homo sapiens cDNA, mRNA sequence.
VERSION     AM955633
XREF        AM955633.1 GI:8145316
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
REFERENCE   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS     Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 447)
Hegde, P., Qi, R., Abernathy, K., Dharap, S., Gaspar, R., Gay, C., Holt,
I.E., Saeed, A.I., Sharov, V., Lee, N.H., Yeatman, T.J. and
Quackenbush, J.
Assessment of gene expression patterns in a model of colon tumor
metastasis using a 19,200 element cDNA microarray
Unpublished (2000)
Contact: John Quackenbush
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 3528
Fax: 301 838 0208
Email: johng@tigr.org
Plate: 81
Seq primer: Reverse.
Location/Qualifiers
1. .447
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="MAGE resequences, MAGD"
/notes="Vector: pBluescriptSKm"

BASE COUNT      184 a      69 c      64 g      130 t
ORIGIN

Query Match      80.6%; Score 87; DB 10; Length 447;
Best Local Similarity 99.0%; Pred. No. 0.24;
Matches 98; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 6 ttgtatattgagatgtttaataaa-tgtgaaaaaatgaataaagcatgttggct 64
|||||
Db 180 TTGTATATGTGAGATGTTTAAATTAATGTGAAAAAATGAATTAAGCATGTTGGTT 239

QY 65 ttccaaagaagaaaaaagaaaaaagaaaaaagaaaaa 103
|||||
Db 240 TTCCAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAA 278

```

RESULT 4
LOCUS N68114/c
DEFINITION yz54a12.s1 Morton fetal Cochlea Homo sapiens CDNA clone
IMAGE:286646.3' similar to gb:J03464 PROCOLLAGEN ALPHA 2(1) CHAIN
PRECUSOR (HUMAN);, mRNA sequence.
ACCESSION N68114
VERSION N68114
KEYWORDS GI:1224275
SOURCE EST.
ORGANISM Homo sapiens
human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 409)
AUTHORS Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiapelli, B.,
Chissee, S., Dietrich, N., Dubuque, T., Favello, A., Gish, W., Hawkins,
M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N., Mardis, E., Moore,
B., Morris, M., Parsons, J., Prange, C., Rifkin, L., Rohlfing, T.,
Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J., Trevaskis, E.,
Underwood, K., Wohlmann, P., Waterston, R., Wilson, R. and Marra, M.,
Generation and analysis of 280,000 human expressed sequence tags
Genome Res. 6 (9), 807-828 (1996)
97044478
TITLE JOURNAL
MEDLINE
COMMENT Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: m13 -40 forward
High quality sequence stop: 291.
FEATURES
source
1. 409
/organism="Homo sapiens"
/db_xref="GDB:3893963"
/db_xref="taxon:9606"
/clone="IMAGE:286846"
/clone_lib="Morton Fetal Cochlea"
/tissue_type="Cochlea"
/dev_stage="16-22 week fetus"
/lab_host="SOLR cells (kanamycin resistant)"
/note="Organ: ear; Vector: pBluescript SK-; Site: 1: EcoRI;
Site: 2: XhoI; Reference: Genomics 23, 42-50 (1994) Cloned
unidirectionally. Primer: Oligo dt. Fetal cochlea, normal.
37% of inserts <0.5 kb, 56% 0.5-1.0 kb, 7% >1 kb. Uni-ZAP
XR Vector. Library constructed by N. Robertson, C. Morton.
-5' adaptor sequence: 5' GAATTCGCGACGAG 3' -3' adaptor
sequence: 5' CTCGAGTGTGTGTGTGTGTGTGTGT 3'

BASE COUNT 143 a 61 c 49 g 155 t 1 others
ORIGIN

Query Match 79.1%, Score 85.4; DB 11; Length 409;
Best Local Similarity 98.0%; Pred. No. 0.38;
Matches 97; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

OY 6 ttgatatgagatgtttaataaa-tgtgaataaataaataaagcattgtgt 64
|||||
DB 107 TTTGATATGTGAGATGTTTAATTAATTGTCAAAAAATGAATTAAGCATGTTGTT 48
|||||
OY 65 ttccaaagaagaaaaaagaaaaaagaaaaaagaaaaa 103
|||||
DB 47 TTCACAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAA 9
|||||

RESULT 5
LOCUS AM069257/c
DEFINITION cr42c12.x1 Jia bone marrow stroma Homo sapiens CDNA clone

ACCESSION HBMS-Cr42c12.3', mRNA sequence.
VERSION AM069257
KEYWORDS AM069257.1 GI:6024178
SOURCE EST.
ORGANISM Homo sapiens
human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 591)
AUTHORS Jia, L.B., Young, M.F., Touchman, J.W., Bouffard, G.G.,
Beckstrom, Sternberg, S.M., Green, E.D., Powell, J.I., Yang, L.M., Robey,
P.G., Hotchkiss, R.N. and Francmanno, C.A.
SCAP: The Skeletal Genome Anatomy Project
Unpublished (1997)
CONTACT: Libin Jia
Medical Genetics Branch
National Human Genome Research Institute
10/10C101, 9000 Rockville Pike, Bethesda, MD 20892-1267, USA
Tel: 301-402-4877
Fax: 301-496-7157
Email: libin@helix.nih.gov
DNA Sequencing and analyses by National Institutes of Health
Intramural Sequencing Center (NISC).
Plate: 42 row: C column: 12
Seq primer: -21M13 forward primer (ABI).
FEATURES
source
1. 591
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="HBMSC-Cr42c12"
/clone_lib="Jia bone marrow stroma"
/sex="mixed"
/tissue_type="bone marrow stroma"
/dev_stage="mixed"
/lab_host="XLI-Blue MRF"/SOLR"
/note="Vector: pBluescript; Site: 1: EcoRI; Site: 2: XhoI;
oligo-dt priming. Directionally cloned. Size-selected for
average insert size >0.5 kb. Library supplied by Dr. Libin
Jia (NHCRI)."
BASE COUNT 192 a 85 c 86 g 209 t 19 others
ORIGIN

Query Match 79.1%, Score 85.4; DB 10; Length 591;
Best Local Similarity 98.0%; Pred. No. 0.29;
Matches 97; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

OY 6 ttgatatgagatgtttaataaa-tgtgaataaataaataaagcattgtgt 64
|||||
DB 99 TTTGATATGTGAGATGTTTAATTAATTGTCAAAAAATGAATTAAGCATGTTGTT 40
|||||
OY 65 ttccaaagaagaaaaaagaaaaaagaaaaaagaaaaa 103
|||||
DB 39 TTCACAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAA 1
|||||

RESULT 6
LOCUS AM022830
DEFINITION df44f07.y1 Morton Fetal Cochlea Homo sapiens CDNA clone
IMAGE:2486557.5', mRNA sequence.
ACCESSION AM022830
VERSION AM022830.1 GI:5876360
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 452)
AUTHORS Robertson, N.G., Khetarpal, U., Gutierrez-Espeleta, G.A., Bieder, F.R.
and Morton, C.C.
Isolation of novel and known genes from a human fetal cochlear CDNA

JOURNAL MEDLINE
 95130111
 COMMENT
 library using subtractive hybridization and differential screening
 Genomics 23, 42-50 (1994)

Contact: Morton, C. C.
 Departments of Pathology and Obstetrics, Gynecology and
 Reproductive Biology
 Brigham and Women's Hospital
 75 Francis Street, Harvard Medical School, Boston, MA 02115, USA
 Tel: 617 732 7980
 Fax: 617 738 6996
 Email: cmorton@rics.bwh.harvard.edu

DNA sequencing and analyses were performed by National Institutes
 of Health Intramural Sequencing Center. (NISC; see
<http://www.nisc.nih.gov>).

This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.

Plate: LLAM6186 row: L column: 14
 Seq primer: M13RPL reverse primer (ABI).

FEATURES

Location/Qualifiers

1..452
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:2486557"
 /clone_1ib="Morton Fetal Cochlea"
 /tissue_type="cochlea"
 /dev_stage="16-22 week fetus"
 /lab_host="SOLR cells (kanamycin resistant)"
 /note="Organ: ear; Vector: pBluescript SK-; Site: 1: EcoRI;
 Site 2: XhoI; Reference: Genomics 23, 42-50 (1994) Cloned
 unidirectionally. Primer: Oligo dT. Fetal cochlea, normal.
 378 of inserts <0.5 kb, 568 0.5-1.0 kb, 78 >1 kb. Unl-ZAP
 XR Vector. Library constructed by N. Robertson, C. Morton.
 -5' adaptor sequence: 5' GAATTCGCGACGAG 3' -3' adaptor
 sequence: 5' CTCGAGTCTTTTCTTTTCTTTT 3' "

BASE COUNT 151 a 61 c 84 g 155 t 1 others

ORIGIN

Query Match 78.3%; Score 84.6; DB 10; Length 452;
 Best Local Similarity 94.2%; Pred. No. 0.43;
 Matches 98; Conservative 0; Mismatches 5; Indels 1; Gaps 1;

QY 6 ttgtatattgagatgttcaataaa-tgtgaaaaaaatgaataagatgttggt 64
 |||||||
 Db 338 TTGTATATGTGAGATGTTAAATTAATTTGAAAAAATAAGCATGTTGGTT 397
 |||||||
 QY 65 ttccaaaagaaaaaataaaaaaataaaaaaagcgcg 108
 |||||||
 Db 398 TTCCAAANGAAAAAATAAAAAAATAAACTCGGG 441
 |||||||

RESULT 7
 AI659779 184 bp mRNA EST 10-MAY-1999
 LOCUS AI659779/c
 DEFINITION t86c08.x1 NCI-CGAP_Pr28 Homo sapiens cDNA clone IMAGE:2248430 3'
 similar to gb:03464 PROCOLLAGEN ALPHA 2(1) CHAIN PRECURSOR (HUMAN
);, mRNA sequence.

ACCESSION AI659779
 VERSION AI659779.1
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 184)
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 Unpublished (1997)
 CONTACT: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.
 Emmert-Buck, M.D., Ph.D.

FEATURES

CDNA Library Preparation: M. Bento Soares, Ph.D.
 CDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/dbfp/image/image.html
 Seq primer: -40UP from Gibco.
 Location/Qualifiers

1..184
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:2248430"
 /clone_1ib="NCI-CGAP_Pr28"
 /sex="male"
 /dev_stage="adult"
 /lab_host="DH10B"
 /note="Organ: prostate; Vector: pT73D-Pac (Pharmacia)
 with a modified polylinker; Plasmid DNA from the
 normalized library NCI-CGAP_Pr22 was prepared, and ss
 circles were made in vitro. Following HAP purification,
 this DNA was used as tracer in a subtractive hybridization
 reaction. The driver was PCR-amplified cDNAs from a pool
 of 5,000 clones made from the same library (clones
 985608-986759, 1101192-1101959, and 1217928-1220615).
 Subtraction by Bento Soares and M. Fatima Bonaldo."

BASE COUNT

57 a 27 c 20 g 80 t

ORIGIN

Query Match 78.1%; Score 84.4; DB 10; Length 184;
 Best Local Similarity 98.0%; Pred. No. 0.85;
 Matches 96; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 6 ttgtatattgagatgttcaataaa-tgtgaaaaaaatgaataagatgttggt 64
 |||||||
 Db 98 TTGTATATGTGAGATGTTAAATTAATTTGAAAAAATAAGCATGTTGGTT 39
 |||||||
 QY 65 ttccaaaagaaaaaataaaaaaataaaaaaagcgcg 102
 |||||||
 Db 38 TTCCAAANGAAAAAATAAAAAAATAAACTCGGG 441
 |||||||

RESULT 8
 AW594372 433 bp mRNA EST 22-MAR-2000
 LOCUS AW594372/c
 DEFINITION hg68d07.x1 NCI-CGAP_G6 Homo sapiens cDNA clone IMAGE:2950765 3'
 similar to gb:03464 PROCOLLAGEN ALPHA 2(1) CHAIN PRECURSOR (HUMAN
);, mRNA sequence.

ACCESSION AW594372
 VERSION AW594372.1
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 433)
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 Unpublished (1997)
 CONTACT: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
 R. Emmert-Buck, M.D., Ph.D.

CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
 Bonaldo, Ph.D.
 CDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
image.llnl.gov/image/html/iresources.shtml

Possible reversed clone: polyT not found

Seq primer: -400P from Gibco
High quality sequence stop: 423.

FEATURES
source

Location/Qualifiers
1. 433
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="2950765"
/clone_lib="NCI-CGAP_GC6"
/tissue_type="pooled germ cell tumors"
/lab_host="DH10B"

BASE COUNT

159 a 71 c 51 g 152 t

ORIGIN

Query Match

Best Local Similarity 97.0%; Pred. No. 0.54; Length 433;

Matches 96; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

6 ttgtatattgtgagatgtttaataa-tgtgaaaaaataagaacatgttgggt 64

107 TTTGTATATGTCAGATGTTTAAATTAATGTAATAAATGAATGAATGTTGGTT 48

65 ttccaaagaaaaaataaaaaaataaaaaa 103

47 TTCCAAAAGAAAAAATAAATAAATAAAGAAAA 9

RESULT 9

LOCUS

A1445464 543 bp mRNA EST 13-APR-1999
tj34h05.x1 NCI-CGAP_Panl Homo sapiens cDNA clone IMAGE:2143449 3'
similar to gp:j03464 PROCOLLAGEN ALPHA 2(1) CHAIN PRECURSOR (HUMAN
); mRNA sequence.

ACCESSION

A1445464

VERSION

A1445464.1 GI:4288395

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

REFERENCE

1 (bases 1 to 543)

AUTHORS

NCI-CGAP

TITLE

http://www.ncbi.nlm.nih.gov/ncicgap.

JOURNAL

unpublished (1997)

COMMENT

Contact: Robert Strausberg, Ph.D.

FEATURES

source

1. 543

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_image="2143449"

/clone_lib="NCI-CGAP_Panl"

/tissue_type="adenocarcinoma"

/lab_host="DH10B"

/note="Organ: pancreas; Vector: pcMV-SPORT6; Site_1: SalI;

Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.72 kb. Life Technologies catalog #:
11548-013"

BASE COUNT

184 a 89 c 74 g 196 t

ORIGIN

Query Match

Best Local Similarity 97.0%; Pred. No. 0.46; Length 543;

Matches 96; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

6 ttgtatattgtgagatgtttaataa-tgtgaaaaaataagaacatgttgggt 64

103 TTTGTATATGTCAGATGTTTAAATTAATGTAATAAATGAATGAATGTTGGTT 44

65 ttccaaagaaaaaataaaaaaataaaaaa 103

43 TTCCCAAAAAAATAAATAAATAAAGAAAAA 5

RESULT 10

LOCUS

B1339163 275 bp mRNA EST 30-JUL-2001
364014 MARC 2P1G Sus scrofa cDNA 5', mRNA sequence.

Accession B1339163

VERSION

B1339163.1 GI:15032446

KEYWORDS

EST.

SOURCE

pig.

ORGANISM

Sus scrofa

REFERENCE

1 (bases 1 to 275)

AUTHORS

Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.

TITLE

Design and use of two pooled tissue normalized cDNA libraries for

EST discovery in swine

unpublished (2000)

Contact: Smith TPL

USDA, ARS, US Meat Animal Research Center

PO Box 166, Clay Center, NE 68933-0166, USA

Tel: 402 762 4366

Fax: 402 762 4390

Email: smith@email.marc.usda.gov

Single pass sequencing. Bases called and alt-trimmed with phred

v0.980904.e. Vector identified by cross-match with the -minscore 18

and -mismatch 12 options.

PCR Primers

FORWARD: AGGAACAGCTATGACCAT

BACKWARD: GTTTCCACGTCAGCAGC

Plate: 100 row: 0 column: 15

Seq primer: ATTAGGTGACATATAG.

FEATURES

source

1. 275

/organism="Sus scrofa"

/db_xref="taxon:9823"

/clone_image="MARC 2P1G"

/tissue_type="pooled"

/lab_host="DH10B"

/note="Vector: pcMV SPORT6; Site_1: XbaI; Site_2: XhoI;

Library made from pooled tissue from testis, ovary,

endometrium, hypothalamus, pituitary, and placenta."

BASE COUNT

104 a 47 c 43 g 80 t

ORIGIN

Query Match

Best Local Similarity 93.3%; Pred. No. 0.78; Length 275;

Matches 98; Conservative 0; Mismatches 5; Indels 2; Gaps 1;

6 ttgtatattgtgagatgtttaataa-tgtgaaaaaataagaacatgttgggt 63

104 TTTGTATATGTCAGATGTTTAAATTAATGTAATAAATGAATGAATGTTGGTT 163

QY 64 ttccaaagagagagagagagagagagagagcg 108
|||||
Db 164 TTTCAGAGAGAGAGAGAGAGAGAGAGAGG 208

RESULT 11
AI620752/c
LOCUS
DEFINITION
AI620752 510 bp mRNA EST 15-DEC-1999
similar to gb:J03464 PROCOLLAGEN ALPHA 2(1) CHAIN PRECURSOR (HUMAN
); mRNA sequence.
AI620752
VERSION
KEYWORDS
SOURCE
ORGANISM
human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS
TITLE
Natl Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL
COMMENT
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bdrp/image/image.html
Insert length: 1768 Std Error: 0.00
Seq primer: -400P from Gibco
High quality sequence stop: 303
POLYA-No.

FEATURES
source
Location/Qualifiers
1. 510
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2258927"
/clone_1b="NCI-CGAP-Gas4"
/tissue_type="poorly differentiated adenocarcinoma with
signet ring cell features"
/lab_host="DH10B"
/note="Organ: stomach; Vector: pCMV-SPORT6; Site_1: Salt;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.69 kb. Life Technologies catalog #:
11549-011"

BASE COUNT 170 a 90 c 73 g 177 t

ORIGIN

Query Match 76.9%; Score 83; DB 10; Length 510;
Best Local Similarity 94.2%; Pred. No. 0.58;
Matches 97; Conservative 0; Mismatches 5; Indels 1; Gaps 1;

QY 2 ccatttgtatgtgagatgttaataa-tgtgaataatgaataagcatgtt 60
|||||
Db 111 CTTTGTGATGATGAGATGTTAAATTAATCTGAAAAAATGAATTAACGATGTT 52

QY 61 ggtttccaaagagagagagagagagagagagagag 103
|||||
Db 51 GGTTCAGAGAGAGAGAGAGAGAGAGAGAGG 9

RESULT 12
AI435232/c
LOCUS
DEFINITION
AI435232 520 bp mRNA EST 30-MAR-1999
similar to gb:J03464 PROCOLLAGEN ALPHA 2(1) CHAIN PRECURSOR (HUMAN
); mRNA sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS
TITLE
Natl Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL
COMMENT
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov
Life Technologies catalog #: 11548-013
DNA Sequencing by: Washington University Genome Sequencing Center
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bdrp/image/image.html
Insert length: 3135 Std Error: 0.00
Seq primer: -400P from Gibco
High quality sequence stop: 280.

FEATURES
source
Location/Qualifiers
1. 520
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2130165"
/clone_1b="NCI-CGAP-Pan1"
/tissue_type="adenocarcinoma"
/lab_host="DH10B"
/note="Organ: pancreas; Vector: pCMV-SPORT6; Site_1: Salt;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.72 kb. Life Technologies catalog #:
11548-013"

BASE COUNT 177 a 90 c 81 g 172 t

ORIGIN

Query Match 76.3%; Score 82.4; DB 10; Length 520;
Best Local Similarity 97.9%; Pred. No. 0.66;
Matches 94; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 6 ttctatcgtgagatgttcaataa-tgtgaataatgaataagcatgtt 64
|||||
Db 96 TTCTATATGTGAGATGTTAAATTAATCTGAAAAAATGAATTAACGATGTTGTT 37

QY 65 ttccaaagagagagagagagagagagagagag 100
|||||
Db 36 TTCCAGAGAGAGAGAGAGAGAGAGAGAGG 1

RESULT 13
AI697004/c
LOCUS
DEFINITION
AI697004 529 bp mRNA EST 17-DEC-1999
similar to gb:J03464 PROCOLLAGEN ALPHA 2(1) CHAIN PRECURSOR (HUMAN
); mRNA sequence.
AI697004
VERSION
KEYWORDS
SOURCE
ORGANISM
human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS
TITLE
Natl Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL
COMMENT
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov
Life Technologies catalog #: 11548-013

DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLMD at:
www.bio.llnl.gov/dbp/image/image.html

Insert Length: 774 Std Error: 0.00

Seq primer: -40UP from Gidco

High quality sequence stop: 284.

FEATURES

SOURCE

1..529

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_image="2324602"

/clone_lib="NCI CGAP Panel"

/tissue_type="adenocarcinoma"

/lab_host="DH10B"

/note="Organ: pancreas; Vector: pCMV-SPORT6; Site_1: SalI;

Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.

Average insert size 1.72 kb. Life Technologies catalog #:

11548-013"

BASE COUNT

171 a 93 c 78 g 187 t

ORIGIN

Query Match 76.1%; Score 82.2; DB 10; Length 529;

Best Local Similarity 96.0%; Pred. No. 0.69;

Matches 95; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

OY 6 ttgtatattgtgagatgtttaataaa-tgtgaaaaaatgaataaagcatgttggtt 64

Db 102 TTTGATATGTCAGATGTTTAAATGTCGAAAAAATGAATGAATGATGTTGGTT 43

OY 65 ttccaaagaaaaaataaaaaaataaaaaaataaaaaa 103

Db 42 TTCCCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4

RESULT 14

LOCUS

AM958385 529 bp mRNA EST 01-JUN-2000

DEFINITION EST370455 MAGE resequences, MAGE Homo sapiens cDNA, mRNA sequence.

ACCESSION AM958385

VERSION AM958385.1 GI:8148069

KEYWORDS

SOURCE

ORGANISM

human.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 529)

Hegde, P., Qi, R., Abernathy, K., Dharap, S., Gaspard, R., Gay, C., Holt

, I.E., Saeed, A.I., Sharov, V., Lee, N.H., Yeatman, T.J. and

Quackenbush, J.

Assessment of gene expression patterns in a model of colon tumor

metastasis using a 19,200 element cDNA microarray

Unpublished (2000)

Contact: John Quackenbush

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 3528

Fax: 301 838 0208

Email: johnq@tigr.org

Plate: 119

Seq primer: Reverse.

Location/Qualifiers

1..529

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_lib="MAGE resequences, MAGE"

/note="Vector: pBluescriptSkm"

BASE COUNT 178 a 79 c 90 g 173 t 9 others

ORIGIN

Query Match 75.9%; Score 82; DB 10; Length 529;

Best Local Similarity 95.9%; Pred. No. 0.72;

Matches 93; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

OY 6 ttgtatattgtgagatgtttaataaa-tgtgaaaaaatgaataaagcatgttggtt 64

Db 317 TTTGATATGTCAGATGTTTAAATGTCGAAAAAATGAATGAATGATGTTGGTT 376

OY 65 ttccaaagaaaaaataaaaaaataaaaaaataaaaaa 101

Db 377 TTCCAAANGAAAAAAAAAAAAAAAAAAAAAAAAA 413

RESULT 15

LOCUS

AU120039 547 bp mRNA EST 19-OCT-2000

DEFINITION AU120039 HEMBA1 Homo sapiens cDNA clone HEMBA1007207 5', mRNA

sequence.

ACCESSION AU120039

VERSION AU120039.1 GI:10935274

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 547)

Ota, T., Nishikawa, T., Suzuki, Y., Ishii, S., Salto, K., Kawai, Y.,

Yamamoto, J., Wakamatsu, A., Nakamura, Y., Nagai, T., Sugano, S. and

Isogai, T.

HRI human cDNA project

Unpublished (2000)

Contact: Takao Isogai

Genomics Laboratory

Helix Research Institute

1532-3 Yana, Kisarazu, Chiba 292-0812, Japan

Tel: 81-438-52-3951

Fax: 81-438-52-3952

Email: genomics@hri.co.jp

HRI human cDNA project: 5'- & 3'-end one pass sequencing. Helix

Research Institute; cDNA library construction; Department of

Virology, Institute of Medical Science, University of Tokyo, and

Helix Research Institute.

Location/Qualifiers

1..547

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_image="HEMBA1007207"

/clone_lib="HEMBA1"

/tissue_type="whole embryo, mainly head"

/dev_stage="embryo, 10 weeks"

/note="Vector: pMR18SFL3"

BASE COUNT 180 a 84 c 93 g 185 t 5 others

ORIGIN

Query Match 75.9%; Score 82; DB 10; Length 547;

Best Local Similarity 97.9%; Pred. No. 0.7;

Matches 93; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

OY 6 ttgtatattgtgagatgtttaataaa-tgtgaaaaaatgaataaagcatgttggtt 64

Db 287 TTTGATATGTCAGATGTTTAAATGTCGAAAAAATGAATGAATGATGTTGGTT 346

OY 65 ttccaaagaaaaaataaaaaaataaaaaaataaaaaa 99

Db 347 TTCCAAAGAAAAAAAAAAAAAAAAAAAAAAAAA 381

Search completed: January 17, 2002, 06:42:50
Job time: 3077 sec

```

XX 18-APR-1996; 96US-0634546.
XX (TEXA ) UNIV TEXAS SYSTEM.
XX
XX PI Singh P, Wood TG;
XX DR MPI, 2001-136591/14.
XX DR P-PSDB; AAY72384.
XX
PT Isolated antisense polynucleotide genetic construct for treatment of
PT colon cancer -
XX
PS Claim 1; Column 41-42; 35pp; English.
XX
CC The present invention relates to methods for the treatment of colon
CC cancer. The method involves inhibiting gastrin expression in
CC colon cancer cells using antisense construct.
CC The present sequence is a human preprogastrin cDNA used in the
CC construction of antisense construct. Gastrin is a peptide hormone
CC that plays a role in the initiation of colon tumors.
XX
SO Sequence 383 BP; 89 A; 120 C; 111 G; 63 T; 0 other;

Query Match
Best Local Similarity 50.1%; Score 52.6; DB 22; Length 383;
Matches 72; Conservative 0; Mismatches 14; Indels 1; Gaps 1;

QY 19 ccggcagctctgtgtcgaatgacgacgacgtgtgtatgtctgtcttgcact 78
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 24 cctccacgactctgcaagacgacgacgacgtgtgtatgtctgtatcttgcact 83
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 79 aggatacagccgacctctctgaagctt 105
   || || || || || || || || || || || || || || || || || || ||
DB 84 -ggctctggccgacctctctgaagctt 109

RESULT 2
ID AAN90234 standard; DNA; 829 BP.
XX
AC AAN90234;
XX
DT 01-NOV-1989 (first entry)
XX
DE Gastrin recombinant DNA.
XX
KW Recombinant DNA; promoter; signal peptide; gastrin; shuttle
   vector; yeast; Escherichia coli.
XX
OS Synthetic.
XX
FH Key
FT exon Location/Qualifiers
FT intron 1..472 /*tag= a
FT intron 473..602 /*tag= b
FT exon 603..829 /*tag= c
FT misc_feature 78..81 /*tag= c
FT misc_feature 288..291 /*tag= d
FT misc_feature 320..325 /*tag= e
FT misc_feature 320..325 /*tag= f
FT misc_feature 436..441 /*tag= g
FT misc_feature 711..716 /*tag= h
FT misc_feature 768..771 /*tag= i
FT misc_feature 794..797 /*tag= j

```

```

FT /*tag= j
FT CDS 262..694
FT /*tag= k
XX
XX JP64002579-A.
XX
XX 03-MAY-1989.
XX
XX 24-JUN-1987; 87JP-0156745.
XX
XX 24-JUN-1987; 87JP-0156745.
XX
XX (AJIN ) AJINOMOTO KK.
XX
XX MPI; 1989-176418/24.
XX
XX P-PSDB; AAP90331.
XX
PT Recombinant DNA encoding gastrin - also comprising promoter and signal
PT peptide coding regions, inserted into shuttle vector to produce
PT gastrin in yeast or E.coli.
XX
PS Claim; page 524 and fig 2; 10pp; Japanese.
XX
CC The recombinant DNA effectively produces gastrin in high quantities.
CC It is made by integrating DNA fragment encoding gastrin into a shuttle
CC vector which can grow in yeast or E.coli, eg pAM82, CY37, pAT77, YEP51,
CC YEP52 or pJB218. The DNA encoding signal peptide is, eg from human
CC gastrin, Mfalpahal (yeast secretion hormone), yeast invertase, killer
CC toxin, yeast alkaline or acid phosphatase, human lysozyme, human alpha-
CC amylase or human trypsin inhibitor. DNA encoding gastrin is from, eg
CC pIG53. See also AAP90331. Misc. feature d is Sau3A site, e is Sau3A
CC site, f is HindIII site, g is PvuII site, h is HindIII site, i is Sau3A
CC site, and j is Sau3A site. The CDS is minus the intron.
XX
SO Bases 262-318 encode signal peptide, and bases 319-429 encode gastrin.

Query Match
Best Local Similarity 92.4%; Score 47.6; DB 10; Length 829;
Matches 61; Conservative 0; Mismatches 4; Indels 1; Gaps 1;

QY 40 gatgcagcagcagatgtgtatgtatgtcgtcttgcactagacgacgacctctcg 99
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 261 gatgcagcagcagatgtgtatgtatgtcgtcttgcact-ggctctggccgacctctcg 319
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 100 aagctt 105
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 320 aagctt 325

RESULT 3
ID AAT97434 standard; cDNA; 613 BP.
XX
AC AAT97434;
XX
DT 05-MAY-1998 (first entry)
XX
DE Human gastrin gene.
XX
KW Human; gastrin; inhibition; colon cancer; colorectal cancer;
   treatment; prophylaxis; ss.
XX
OS Homo sapiens.
XX
FH Key
FT CDS Location/Qualifiers
FT CDS 1..210 /*tag= a
FT /*product= "gastrin"
FT /note= "no stop codon given; if the coding sequence is
   extended to position 225, it ends in a stop
   codon and Gly-Arg-Arg-Cys is added to the end

```

[illegible]

```

FT      /note= "Does not include stop codon"
FT      /partial
PN      US6165990-A.
PD      26-DEC-2000.
PE      15-MAY-1998; 98US-0079372.
PR      18-APR-1996; 96US-0634546.
PA      (TEXA ) UNIV TEXAS SYSTEM.
PI      Singh P, Wood TG;
PI      WPI: 2001-136591/14.
DR      P-PSDB: AAV72377.
PT      Isolated antisense polynucleotide genetic construct for treatment of
PT      colon cancer -
PS      Disclosure: Column 35-36; 35pp: English.
CC      The present invention relates to methods for the treatment of colon
CC      cancer. The method involves inhibiting gastrin expression in
CC      colon cancer cells using antisense construct.
CC      The present sequence is a human partial preprogastrin DNA. Gastrin is
CC      a peptide hormone that plays a role in the initiation of colon tumours.
SQ      Sequence 613 BP; 135 A; 181 C; 173 G; 124 T; 0 other;
SQ
Query Match          44.4%; Score 46.6; DB 22; Length 613;
Best Local Similarity 92.3%; Pred. No. 2.1e-07;
Matches 60; Conservative 0; Mismatches 4; Indels 1; Gaps 1
QY      41 atgcagcagcagtgtgtatgtctgtccttgcactagatcagccgccttctctga 100
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      1 atgcagcactatgtgtgtatgtctgtcacttgcact-ggctctggccgccttctctga 59
QY      101 agcct 105
        ||| ||| ||| |||
DB      60 agcct 64

RESULT 5
AAH18005/c
ID      AAH18005 standard; cDNA; 2857 BP.
XX
XX      AAH18005;
AC
XX
XX      26-JUN-2001 (first entry)
DT
XX
XX      Human cDNA sequence SEQ ID NO:17809.
XX
XX      Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
XX      Homo sapiens.
OS
XX
XX      EP1074617-A2.
PN
XX
XX      07-FEB-2001.
PD
XX
XX      28-JUL-2000; 2000EP-0116126.
PE
XX
XX      29-JUL-1999; 99JP-0248036.
PR
XX      27-AUG-1999; 99JP-0300253.
PR      11-JAN-2000; 2000JP-0118776.
PR      02-MAY-2000; 2000JP-0183767.
PR      09-JUN-2000; 2000JP-0241899.
XX
XX      (HELI-) HELIX RES INST.
XX

```

[illegible]

PR		25-APR--2000;	2000US-0552317.	
PR		09-JUL-2000;	2000US-0598042.	
PR		19-JUL-2000;	2000US-0620312.	
PR		03-AUG-2000;	2000US-0653450.	
PR		14-SEP-2000;	2000US-0662191.	
PR		19-OCT-2000;	2000US-0693036.	
PR		29-NOV-2000;	2000US-0727344.	
XX		(HYSE-) HYSEQ INC.		
PA				
XX	Tang YT,	Liu C,	Ashundi V,	Chen R, Ma Y, Qian XB, Ren F, Wang D;
Pt	Wang Z,	Wehman T,	Xu G,	Xue AJ, Yang Y, Zhang J;
Pt	Zhao QA,	Zhou P,	Goodrich R,	Dzmanac RT;
DR	WI:	2001-442253/47.		
DR	P-PSTB:	AAM39311.		
XX				
PT		Novel nucleic acids and polypeptides, useful for treating disorders such as central nervous system injuries -		
XX				
PS	Claim 1;	SEQ ID NO 670; 10078pp;	English.	
CC		The invention relates to human nucleic acids (AAI57798-AAI61369) and the encoded polypeptides (AAM38642-AAM42213) with nootropic,		
CC		Immunosuppressant and cyostatic activity. The polynucleotides are useful in gene therapy. A composition containing a polypeptide or polynucleotide of the invention may be used to treat diseases of the peripheral nervous system, such as peripheral nervous injuries, peripheral neuropathy and localised neuropathies and central nervous system diseases, such as Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis, and Shy-Drager Syndrome. Other uses include the utilisation of the activities such as: immune system suppression, CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic and thrombolytic activity, cancer diagnosis and therapy, drug screening, CC assays for receptor activity, arthritis and inflammation, leukemias and CC C.N.S disorders.		
CC	Note:	The sequence data for this patent did not form part of the printed specification.		
XX				
SQ	Sequence	4495 BP; 1401 A; 945 C; 1040 G; 1109 T; 0 other;		
OY	Query Match	28.8%; Score 30.2; DB 22; Length 4495;		
	Best Local Similarity .69.5%;	Pred.No. 0.35;		
Matches	41; Conservative	0; Mismatches 18; Indels 0; Gaps 0		
Df	3983 ACATCGTGGAGTTCTGTCGTACAGATTACTCAAGAAGTGCCTGGTGCTTTGTTCTTT	3925		
RESULT 7				
AI60253/c				
ID AI60253 standard;	cDNA: 4775 BP.			
XX	AAI60253;			
XX				
DT	22-OCT-2001 (first entry)			
DE	Human polynucleotide SEQ ID NO 4242.			
KW	nootropic; immunosuppressant; cytostatic; gene therapy; cancer; peripheral nervous system; neuropathy; central nervous system; CNS; Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic; amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic; chemokinetic; thrombolytic; drug screening; arthritis; inflammation; leukaemia; ss.			
OS	Homo sapiens.			
NN	WO200153312-A1.			
DP	26-JUL-2001			

XX	26-DEC-2000; 2000WO-US34263.
PR	21-JAN-2000; 2000US-0488725.
PR	25-APR-2000; 2000US-0552317.
PR	09-JUL-2000; 2000US-0596042.
PR	19-JUL-2000; 2000US-0620312.
PR	03-AUG-2000; 2000US-0633450.
PR	14-SEP-2000; 2000US-0662191.
PR	19-OCT-2000; 2000US-0693036.
PR	29-NOV-2000; 2000US-0727344.
PA	(HYSE-) HYSEQ INC.
P1	Tang Y T, Liu C, Asundi V, Chen R, Ma Y, Qian X B, Ren F, Wang D;
P1	Wang J, Wang Z, Wehrman T, Xu C, Xue A U, Yang Y, Zhang J;
P1	Zhao Q A, Zhou P, Goodrich R, Drmanac R T;
XX	WPI: 2001-442253/47.
DR	P-PSDB; AAM41097.
PT	Novel nucleic acids and polypeptides, useful for treating disorders
PT	such as central nervous system injuries -
XX	Claim 1; SEQ ID NO 4242; 10076bp; English.
PS	The invention relates to human nucleic acids (AAI57798-AAI61369) and
XX	the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC	immunosuppressant and cytostatic activity. The polynucleotides are useful
CC	in gene therapy. A composition containing a polypeptide or polynucleotide
CC	of the invention may be used to treat diseases of the peripheral nervous
CC	system, such as peripheral nervous injuries, peripheral neuropathy and
CC	localised neuropathies and central nervous system diseases, such as
CC	Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC	lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC	utilisation of the activities such as: immune system suppression,
CC	Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC	and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC	assays for receptor activity, arthritis and inflammation, leukaemias and
CC	C.N.S disorders.
CC	Note: The sequence data for this patent did not form part of the printed
CC	specification.
SQ	Sequence 4775 BP; 1494 A; 999 C; 1122 G; 1160 T; 0 other;
OY	Query Match 28.8%; Score 30.2; DB 22; Length 4775;
OY	Best Local Similarity 69.5%; Pred. No. 0.36;
Matches	41; Conservative 0; Mismatches 18; Indels 0; Gaps 0
Df	15 acatccgcgaattctgtgtagcatgatcagcgagactgtgtgtatgtgcctcttt 73 3944 ACTTCTGTGACTTCTCTCTCAGATACACTCAAGAAGCTGGCTGTGGTGTTGTCTTT 3886
RESULT	8
ID	AAH41086/C
XX	AAH41086 standard; CDNA: 6006 BP.
XX	AAH41086;
DT	29-AUG-2001 (first entry)
DE	Human calcium binding/actin crosslinking protein (CBACP) cDNA.
KW	Human; calcium-binding/actin cross-linking protein; CBACP; infertility;
KW	immunological disease; hypertension; renal insufficiency; infection;
KW	muscular dystrophy; malnutrition; neurotransmission disorder; cardiac;
KW	kidney disease; anaemia; heart disease; cancer; immunomodulatory;
KW	hypertensive; nephritic; antiviral; antibacterial; antihaemic;
KW	cystostatic; ss.
XX	Homo sapiens
DS	

[illegible]

XX 25-FEB-2000; 2000EP-0301439.
PF
XX
PR 25-FEB-1999; 990S-0121825.
PR 05-MAR-1999; 990S-0123180.
PR 09-MAR-1999; 990S-0123548.
PR 23-MAR-1999; 990S-0125788.
PR 25-MAR-1999; 990S-0126264.
PR 29-MAR-1999; 990S-0126785.
PR 01-APR-1999; 990S-0127462.
PR 06-APR-1999; 990S-0128234.
PR 16-APR-1999; 990S-0129845.
PR 19-APR-1999; 990S-0130077.
PR 21-APR-1999; 990S-0130449.
PR 23-APR-1999; 990S-0130510.
PR 28-APR-1999; 990S-0130891.
PR 30-APR-1999; 990S-0131449.
PR 30-APR-1999; 990S-0132048.
PR 04-MAY-1999; 990S-0132407.
PR 05-MAY-1999; 990S-0132484.
PR 06-MAY-1999; 990S-0132486.
PR 07-MAY-1999; 990S-0132487.
PR 11-MAY-1999; 990S-0132863.
PR 14-MAY-1999; 990S-0134256.
PR 14-MAY-1999; 990S-0134218.
PR 14-MAY-1999; 990S-0134219.
PR 14-MAY-1999; 990S-0134221.
PR 18-MAY-1999; 990S-0134370.
PR 19-MAY-1999; 990S-0134768.
PR 20-MAY-1999; 990S-0134941.
PR 21-MAY-1999; 990S-0135124.
PR 24-MAY-1999; 990S-0135353.
PR 25-MAY-1999; 990S-0135629.
PR 27-MAY-1999; 990S-0136392.
PR 28-MAY-1999; 990S-0136782.
PR 01-JUN-1999; 990S-0137222.
PR 03-JUN-1999; 990S-0137528.
PR 04-JUN-1999; 990S-0137502.
PR 07-JUN-1999; 990S-0137724.
PR 08-JUN-1999; 990S-0138094.
PR 10-JUN-1999; 990S-0138540.
PR 10-JUN-1999; 990S-0138847.
PR 14-JUN-1999; 990S-0139119.
PR 16-JUN-1999; 990S-0139452.
PR 17-JUN-1999; 990S-0139492.
PR 18-JUN-1999; 990S-0139454.
PR 18-JUN-1999; 990S-0139455.
PR 18-JUN-1999; 990S-0139456.
PR 18-JUN-1999; 990S-0139457.
PR 18-JUN-1999; 990S-0139458.
PR 18-JUN-1999; 990S-0139459.
PR 18-JUN-1999; 990S-0139460.
PR 18-JUN-1999; 990S-0139461.
PR 18-JUN-1999; 990S-0139463.
PR 18-JUN-1999; 990S-0139750.
PR 18-JUN-1999; 990S-0139763.
PR 21-JUN-1999; 990S-0139617.
PR 22-JUN-1999; 990S-0139689.
PR 23-JUN-1999; 990S-0140353.
PR 23-JUN-1999; 990S-0140354.
PR 28-JUN-1999; 990S-0140695.
PR 28-JUN-1999; 990S-0140823.
PR 29-JUN-1999; 990S-0140991.
PR 30-JUN-1999; 990S-0141287.
PR 01-JUL-1999; 990S-0141842.
PR 01-JUL-1999; 990S-0142154.
PR 02-JUL-1999; 990S-0142055.
PR 06-JUL-1999; 990S-0142390.

PR 08-JUL-1999; 990S-0142803.
PR 09-JUL-1999; 990S-0142920.
PR 12-JUL-1999; 990S-0142977.
PR 13-JUL-1999; 990S-0143542.
PR 14-JUL-1999; 990S-0143624.
PR 15-JUL-1999; 990S-0144005.
PR 16-JUL-1999; 990S-0144085.
PR 16-JUL-1999; 990S-0144086.
PR 19-JUL-1999; 990S-0144325.
PR 19-JUL-1999; 990S-0144331.
PR 19-JUL-1999; 990S-0144332.
PR 19-JUL-1999; 990S-0144333.
PR 19-JUL-1999; 990S-0144334.
PR 19-JUL-1999; 990S-0144335.
PR 20-JUL-1999; 990S-0144352.
PR 20-JUL-1999; 990S-0144632.
PR 20-JUL-1999; 990S-0144884.
PR 21-JUL-1999; 990S-0144814.
PR 21-JUL-1999; 990S-0145086.
PR 21-JUL-1999; 990S-0145088.
PR 22-JUL-1999; 990S-0145085.
PR 22-JUL-1999; 990S-0145087.
PR 22-JUL-1999; 990S-0145089.
PR 22-JUL-1999; 990S-0145192.
PR 23-JUL-1999; 990S-0145145.
PR 23-JUL-1999; 990S-0145224.
PR 26-JUL-1999; 990S-0145276.
PR 27-JUL-1999; 990S-0145513.
PR 27-JUL-1999; 990S-0145518.
PR 27-JUL-1999; 990S-0145519.
PR 28-JUL-1999; 990S-0145591.
PR 02-AUG-1999; 990S-0146386.
PR 02-AUG-1999; 990S-0146388.
PR 03-AUG-1999; 990S-0146389.
PR 04-AUG-1999; 990S-0147038.
PR 04-AUG-1999; 990S-0147204.
PR 05-AUG-1999; 990S-0147302.
PR 05-AUG-1999; 990S-0147192.
PR 06-AUG-1999; 990S-0147260.
PR 06-AUG-1999; 990S-0147303.
PR 06-AUG-1999; 990S-0147416.
PR 09-AUG-1999; 990S-0147493.
PR 09-AUG-1999; 990S-0147935.
PR 10-AUG-1999; 990S-0148171.
PR 11-AUG-1999; 990S-0148319.
PR 12-AUG-1999; 990S-0148341.
PR 13-AUG-1999; 990S-0148565.
PR 13-AUG-1999; 990S-0148564.
PR 16-AUG-1999; 990S-0149368.
PR 17-AUG-1999; 990S-0149175.
PR 18-AUG-1999; 990S-0149426.
PR 20-AUG-1999; 990S-0149722.
PR 20-AUG-1999; 990S-0149723.
PR 20-AUG-1999; 990S-0149929.
PR 23-AUG-1999; 990S-0149929.
PR 23-AUG-1999; 990S-0149930.
PR 26-AUG-1999; 990S-0150566.
PR 26-AUG-1999; 990S-0150884.
PR 27-AUG-1999; 990S-0151065.
PR 27-AUG-1999; 990S-0151066.
PR 27-AUG-1999; 990S-0151080.
PR 30-AUG-1999; 990S-0151303.
PR 31-AUG-1999; 990S-0151438.
PR 01-SEP-1999; 990S-0151930.
PR 07-SEP-1999; 990S-0152363.
PR 10-SEP-1999; 990S-0153070.
PR 13-SEP-1999; 990S-0153758.
PR 15-SEP-1999; 990S-0154018.
PR 16-SEP-1999; 990S-0154039.
PR 20-SEP-1999; 990S-0154779.
PR 22-SEP-1999; 990S-0155139.
PR 23-SEP-1999; 990S-0155486.

XX W09713855-A1.
 XX
 XX 17-APR-1997.
 PD
 XX
 XX 10-OCT-1995; 95WO-EP03988.
 PE
 XX 10-OCT-1995; 95WO-EP03988.
 PR
 XX (NOVS) NOVARTIS AG.
 PA
 XX
 PI Pluschke G, Schmid P;
 XX
 XX WPI: 1997-235891/21.
 DR
 XX P-PSDB; AAW15566.
 DR
 XX
 PT Melanoma-associated chondroitin sulphate proteoglycan - useful for
 PT prophylaxis or treatment of melanoma-associated tumours
 XX
 PS
 XX Claim 6; Page 31-48; 88pp; English.
 XX
 CC A cDNA clone (AAT60450) codes for melanoma-associated chondroitin
 CC sulphate proteoglycan (MCSP) (AAW15566). The complete coding
 CC sequence and 3' untranslated region were obtd. by sequencing small
 CC overlapping cDNA clones produced by PCR amplification (see also
 CC AAT60451-83), and by analysis of cDNA clone lambda M3.1 obtd.
 CC from a melanoma cDNA library. MCSP cDNA, esp. sequences excluding
 CC nucleotides 4867-7898 or 4858-5357, can be used to produce
 CC recombinant MCSP in host cells for use in vaccines and in methods
 CC for the treatment and diagnosis of melanoma, sarcoma and
 CC glioblastoma. Probes (AAT60462-67) derived from MCSP cDNA can be
 CC used to identify novel non-human homologues of MCSP and in tumour
 CC diagnosis, e.g. to localise MCSP mRNA in primary melanomas or
 CC metastatic lesions.
 XX
 X0 Sequence 7918 BP; 1451 A; 2551 C; 2455 G; 1460 T; 1 other;

[illegible]

PR	25-APR-2000;	20000US-0552317.	
PR	09-JUL-2000;	20000US-0598042.	
PR	19-JUL-2000;	20000US-0620312.	
PR	03-AUG-2000;	20000US-0634550.	
PR	14-SEP-2000;	20000US-0662191.	
PR	19-OCT-2000;	20000US-0693036.	
PR	29-NOV-2000;	20000US-0727344.	
XX			
PA	(HYSE-) HYSEQ INC.		
PI	Tang YF, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;		
PI	Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;		
PI	Zhao QA, Zhou P, Goodrich R, Drmanac RT;		
XX			
DR	WPI: 2001-442253/47.		
DR	P-PSDB; AAM38733.		
XX			
PT	Novel nucleic acids and polypeptides, useful for treating disorders		
XX	such as central nervous system injuries -		
PS	Claim 1: SEQ ID NO 92; 10078bp; English.		
XX			
CC	The invention relates to human nucleic acids (AA157798-AA161369) and		
CC	the encoded polypeptides (AAM38642-AAM42213) with neurotropic,		
CC	immunosuppressant and cytostatic activity. The polynucleotides are useful		
CC	in gene therapy. A composition containing a polypeptide or polynucleotide		
CC	of the invention may be used to treat diseases of the peripheral nervous		
CC	system, such as peripheral nervous injuries, peripheral neuropathy and		
CC	localised neuropathies and central nervous system diseases, such as		
CC	Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic		
CC	lateral sclerosis, and Shy-Drager Syndrome. Other uses include the		
CC	utilisation of the activities such as: Immune system suppression,		
CC	Activin/Inhibin activity, chemotactic/chemokinetic activity, haemostatic		
CC	and thrombolytic activity, cancer diagnosis and therapy, drug screening,		
CC	assays for receptor activity, arthritis and inflammation, leukaemias and		
CC	C.N.S disorders.		
CC	Note: The sequence data for this patent did not form part of the printed		
CC	specification.		
XX			
SO	Sequence 8570 BP; 1594 A; 2711 C; 2648 G; 1617 T; 0 other;		
XX			
Query Match	25.9%; Score 27.2; DB 22; Length 8570;		
Best Local Similarity	67.9%; Pred. No.5.3;		
Matches 38; Conservative	0; Mismatches 18; Indels 0; Gaps 0;		
QY	44 cagcagcagatgtgtatgtctgtcgtcttgcactagatcagcgcgtctctcg 99		
DB	957 cagtgatgatgtgtgcccgtgcctctctcgtggccaccactctcgtgtgcctccctg 1012		
II			
RESULT 12			
AA159675/c			
ID	AA159675 standard; cDNA: 8689 BP.		
XX			
AC	AA159675;		
XX			
DT	22-OCT-2001 (first entry)		
DE	Human polynucleotide SEQ ID NO 3664.		
XX			
KM	Human; neurotropic; immunosuppressant; cytostatic; gene therapy; cancer;		
KM	peripheral nervous system; neuropathy; central nervous system; CNS;		
KM	Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;		
KM	amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;		
KM	chemokinetic; thrombolytic; drug screening; arthritis; inflammation;		
KM	leukaemia; ss.		
XX			
OS	Homo sapiens.		
XX			
PN	WO200153312-A1.		
XX			
DD	26-JUL-2001.		

XX	26-DEC-2000;	2000MO-US34263.
PR	21-JAN-2000;	2000US-0488725.
PR	25-APR-2000;	2000US-0552317.
PR	09-JUL-2000;	2000US-0598042.
PR	19-JUL-2000;	2000US-0620312.
PR	03-AUG-2000;	2000US-0653450.
PR	14-SEP-2000;	2000US-0662191.
PR	19-OCT-2000;	2000US-0693036.
PR	29-NOV-2000;	2000US-0727344.
PA	(HYSEQ-) HYSEQ INC.	
PI	Tang Y ^T , Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;	
PI	Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;	
PI	Zhao QA, Zhou P, Goodrich R, Drmanac RT;	
DR	WPI: 2001-442253/47.	
DR	P-PSDB: AAM40519.	
PT	Novel nucleic acids and polypeptides, useful for treating disorders	
PT	such as central nervous system injuries -	
XX		
PS	Claim 1; SEQ ID NO 3664; 10078bp; English.	
XX	The invention relates to human nucleic acids (AAI57798-AAI61369) and	
CC	the encoded polypeptides (AAM8642-AAM42213) with nootropic,	
CC	immunosuppressant and cytostatic activity. The polynucleotides are useful	
CC	in gene therapy. A composition containing a polypeptide or polynucleotide	
CC	of the invention may be used to treat diseases of the peripheral nervous	
CC	system, such as peripheral nervous injuries, peripheral neuropathy and	
CC	localised neuropathies and central nervous system diseases, such as	
CC	Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic	
CC	lateral sclerosis, and Shy-Dreger syndrome. Other uses include the	
CC	utilisation of the activities such as: Immune system suppression,	
CC	Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic	
CC	and thrombolytic activity, cancer diagnosis and therapy, drug screening,	
CC	assays for receptor activity, arthritis and inflammation, leukaemia and	
CC	C.N.S disorders.	
CC	Note: The sequence data for this patent did not form part of the printed	
SQ	specification.	
XX		
XX	Sequence 8689 BP; 1638 A; 2669 C; 2755 G; 1627 T; 0 other;	
XX		
Query Match	25.9%; Score 27.2; DB 22; Length 8689;	
Best Local Similarity	67.9%; Pred. No. 5.3;	
Matches 38; Conservative	0; Mismatches 18; Indels 0; Gaps 0;	
OY	44 cagcgcagatgttgtatgtctgctcttgcactagatcagcgccctctctg 99	
Db	7614 CAGTGATGATGTGGCCTTGTGCGCTTCCTGTGGGCCACACTCTGTGCTGCCTCCCTG 7559	
RESULT 13		
AAF27661/C		
ID	AAF27661 standard; DNA: 2404 BP.	
XX		
AC	AAF27661;	
DT	30-MAR-2001 (first entry)	
XX		
DE	DNA encoding human transcriptional regulator protein #30.	
XX		
KM	Human; transcriptional regulator protein; TXRG; ds.	
OS	Homo sapiens.	
XX		
FN	WO200078954-A2.	
XX		
XD	28-DEC-2000.	
XX		

[illegible]

PR	05-APR-1999;	99US-0137728.	
PR	30-MAR-2000;	2000US-0540763.	
XX			
PA	(CURA-) CURAGEN CORP.		
PI	Shimkets RA, Leach M;		
XX			
DR	WPI, 2000-602362/57.		
DR	P-PSDB; AAB42468.		
PT	Novel nucleic acids and peptides derived from open reading frame X,		
PT	useful for treating e.g. cancers, proliferative disorders,		
PT	neurodegenerative disorders and cardiovascular disease -		
PS	Claim 5; Page 3649-3651; 5507pp; English.		
XX			
CC	AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,		
CC	which represent the human ORFX open reading frames 1 to 3161. The ORFX		
CC	sequences have activities such as: cytostatic; hepatotropic; vulnerary;		
CC	antiproliferative; antiparkinsonian; nootropic; neuroprotective;		
CC	osteopathic; anticonvulsant; antiarrhythmic; immunosuppressant;		
CC	immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;		
CC	antidiabetic; hypotensive; dermatological; immunosuppressive;		
CC	antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;		
CC	antihypertoid; and antihaemic. The sequences can be used for determining		
CC	the presence of or predisposition to, or preventing or treating		
CC	pathological conditions associated with an ORFX-associated disorder. The		
CC	nucleic acids can be used to express ORFX proteins in gene therapy		
CC	vectors. The proteins and nucleic acids may be used to treat cancers,		
CC	proliferative disorders, neurodegenerative disorders, osteoarthritis,		
CC	grief vs host disease, cardiovascular disease, diabetes mellitus,		
CC	hyperextension, hypothyroidism, cholesterol ester storage, systemic lupus		
CC	erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,		
CC	bacterial or fungal infection, malaria, autoimmune disorders, asthma,		
CC	allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,		
CC	nocturnal haemoglobinuria, antiinflammatory disease; to enhance		
CC	coagulation; to inhibit thrombosis; and as a contraceptive.		
XX			
S0	Sequence 2662 BP; 576 A; 758 C; 796 G; 530 T; 2 other:		
QY	6 tgttcgcagacatcctgcagctcgtttagatgatgcagcgacagatgctgatatgcgc 65		
Db	1391 TGGTGACACAGCGCGCGTGATGCTCTTGTGACGCGATGGCGAGCGCTGATGTTGATGATGC 1322		
QY	66 tgcctcttcgc 75		
Db	1321 TGGCGTTTGC 1312		
RESULT 15			
AAH12410			
XX	AAH12410 standard; cDNA: 501 BP.		
XX	AAH12410;		
XX	26-JUN-2001 (first entry)		
XX	Human cDNA clone (3'-primer) SEQ ID NO:9245.		
XX	Human; primer: detection; diagnosis; antisense therapy; gene therapy; ss.		
XX	Homo sapiens.		
OS			
XX	EP1074617-A2.		
PN			
DD	07-FEB-2001.		
XX			
PF	28-JUL-2000; 2000EP-0116126.		

```

XX 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
XX WPI; 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
PS Claim 3: SEQ ID 9245; 2537pp + CD ROM; English.
XX
XX The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SQ Sequence 501 BP; 145 A; 95 C; 86 G; 161 T; 14 other;

```

```

Query Match 25.3%; Score 26.6; DB 22; Length 501;
Best Local Similarity 64.4%; Pred. No. 3.7;
Matches 38; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

```

```

OY 15 acatcctgcagctctgtgtaagatgatgagcagatgtgtatgctgctctt 73
   || ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 255 acttcctgtgagttctctgcacanaactgcacaaaacctgcgtgctgtcttctt 313

```

Search completed: January 17, 2002, 06:46:54
Job time: 3282 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic --nucleic search, using sw model

Run on: January 17, 2002, 08:48:23 ; Search time 2206.96 Seconds
(without alignments)
807.308 Million cell updates/sec

Title: US-09-823-101-6
Perfect score: 108
Sequence: 1 acctattgtatattgagag.....aaaaaaaaaaaaaacgcgcg 108

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues
Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_ov:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pt:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_om:*
20: em_or:*
21: em_ov:*
22: em_pat:*
23: em_ph:*
24: em_pl:*
25: em_ro:*
26: em_sts:*
27: em_sy:*
28: em_un:*
29: em_vl:*
30: em_htgo_hum:*
31: em_htgo_inv:*
32: em_htgo_rtd:*
33: em_htg_hum:*
34: em_htg_inv:*
35: em_htg_rtd:*
36: em_htg_other:*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

8

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	75	69.4	343	11 G62185	G62185 EST544 Foet
2	61.8	57.2	1130	9 HUMC1A22	K01078 Human proco
3	61.8	57.2	38682	9 AF004877	AF004877 Homo sapi
4	61.8	57.2	141120	9 AC002528	AC002528 Human BAC
5	61.8	57.2	188781	2 AC022546	AC022546 Homo sapi
6	61.8	57.2	189401	2 AC092511	AC092511 Papio cyn
7	59.8	55.4	846	11 G15893	G15893 human STS C
8	59.8	55.4	5086	9 HSPPA2ICO	Z74616 H. sapiens m
9	59.8	55.4	5416	9 HUMC1A2	J03464 Human COL1A
10	56	51.9	236699	2 AC026891	AC026891 Mus muscu
11	54.2	50.2	1162	8 AF117707	AF117707 Lycopersi
12	54.2	50.2	3265	6 AX092338	AX092338 Sequence
13	53.6	49.6	4284	3 DDU20661	U20661 Dictyosteli
14	53.6	49.6	150844	33 AL357036	AL357036 Human DNA
15	53.6	49.6	169224	2 AC092814	AC092814 Homo sapi
16	53.6	49.6	193570	2 AL592297	AL592297 Homo sapi
17	53	49.1	542	3 PPNRSUB	X80129 P. falciparu
18	52.8	48.9	1052	3 AB070132	AB070132 Macaca fa
19	52.8	48.9	3499	9 HSU80919	HSU80919 Homo sapien
20	52.8	48.9	4678	9 HSU80917	HSU80917 Homo sapien
21	52.6	48.7	1804	5 TNU72071	U72071 Tilapia nil
22	52.6	48.7	2021	10 BC007159	BC007159 Mus muscu
23	52.6	48.7	2758	9 HSM802938	AL583915 Homo sapi
24	52	48.1	682	9 HST000239	AL360193 Homo sapi
25	51.6	47.8	5178	9 HSM802206	AL137716 Homo sapi
26	51.2	47.4	2802	9 AF084644	AF084644 Homo sapi
27	51.2	47.4	2905	9 AF084645	AF084645 Homo sapi
28	51	47.2	905	3 AF081800	AF081800 Dictyoste
29	50.8	47.0	1064	5 AF067856	AF067856 Xenopus l
30	50.8	47.0	1658	10 AB030908	AB030908 Rattus no
31	50.8	47.0	2028	9 AB025432	AB025432 Homo sapi
32	50.6	46.9	2216	9 AK026788	AK026788 Homo sapi
33	50.4	46.7	2600	4 SSJ001202	AJ001202 Sus scrofa
34	50.2	46.5	3277	9 AF102546	AF102546 Homo sapi
35	50.2	46.5	3423	10 BC010717	BC010717 Mus muscu
36	50.2	46.5	234112	3 PPMALAP2	AL035675 Plasmodiu
37	50	46.3	934	3 DDU25660	U25660 Dictyosteli
38	50	46.3	1118	8 ATP22A	Y08781 A. thaliana
39	50	46.3	1310	3 AB013088	AB013088 Holotrich
40	50	46.3	2976	3 DDU09478	U09478 Dictyosteli
41	49.8	46.1	835	11 CNS06JGB	AL401553 T7 end of
42	49.8	46.1	1045	10 BC006754	BC006754 Mus muscu
43	49.8	46.1	3314	3 AF163834	AF163834 Dictyoste
44	49.6	45.9	1724	9 BC003101	BC003101 Homo sapi
45	49.6	45.9	2568	3 DDP8A7	X51947 Dictyosteli

ALIGNMENTS

RESULT 1
LOCUS G62185
DEFINITION EST544 Foetal cDNA library directionally cloned in Unizap XR vector
Equis caballus STS cDNA 3', sequence tagged site.
ACCESSION G62185
VERSION G62185.1 GI:7677541
KEYWORDS STS.
SOURCE horse.
ORGANISM Equus caballus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
REFERENCE 1 (bases 1 to 343)
AUTHORS Brandon, R., Thomas, S. and Bell, K.
TITLE Development of equine expressed sequence tags for use in radiation
hybrid mapping
JOURNAL Unpublished (2000)
COMMENT

Contact: Richard Brandon
Australian Equine Blood Typing Research Laboratory

University of Queensland
St Lucia, Brisbane, Queensland, 4072, Australia
Tel: 61733658501
Fax: 61733654899
Email: brandone@plok.unq.edu.au
Primer A: GAGCATGTGTCGTGCTGAGAG
Primer B: ACCATGGCAAGATGATAAC
STS size: 205
PCR profile:

Cycle 1:

95 degrees C for 3.00 minute
60 degrees C for 1.00 minute
72 degrees C for 1.00 minute

Cycles 2-31:

95 degrees C for 0.45 minute
60 degrees C for 1.00 minute
72 degrees C for 1.00 minute
95 degrees C for 0.45 minute
60 degrees C for 1.00 minute
72 degrees C for 10.00 minute

Cycle 32:

95 degrees C for 0.45 minute
60 degrees C for 1.00 minute
72 degrees C for 10.00 minute

Thermal Cycler: MJ Research PTC-100
Protocol:

Template: 100ng
Primer: each 0.5 uM
dNTPs: each 200 uM
Taq Polymerase: 0.02 units/uM
Annealing Temp: 60
Total Vol: 20 uL

Buffer:

MgCl2: 1.5 mM
(NH4)2SO4: 16.6 mM
Tris-HCl: 67 mM
Gelatin: 200 ug/ml
Triton X-100 0.45%

PCR works in the horse and not mouse for specific use in RH panel screening.

FEATURES

source

Location/Qualifiers
1..343
/organism="Equus caballus"
/db_xref="taxon:9796"
/clone_lib="foetal CDNA library directionally cloned in Unizap XR vector"
/tissue_type="60-day foetus"
/note="Vector: Bluescript; V-type: phagemid; A whole foetus was used as source of RNA in the construction of a stratagene Unizap XR directional library. The phagemid Bluescript was excised from the virus vector and used as template for sequencing reactions from the 3' end of the cDNA."
1..343
/gene="COL1A2"
/note="collagen pro-alpha 2 (I)"
1..205
/gene="COL1A2"
STS
1..24
/gene="COL1A2"
primer_bind
complement(185..205)
BASE COUNT 118 a 48 c 57 g 119 t 1 others
ORIGIN

Query Match 69.4%; Score 75; DB 11; Length 343;

Best Local Similarity 98.9%; Pred. No. 0.0022; Mismatches 0; Indels 1; Gaps 1;

Db 257 TTGTATATGTGAGATCTTAATAATGTGAAAAAAGAAATGAATGAATGATGTTGGTT 316

QY 65 ttccaaagaaaaaataaaaaa 91

Db 317 TTCCAAAAGAAAAA 343

RESULT 2

HUMC1A22

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SEGMENT

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

FEATURES

source

Location/Qualifiers
1..1130
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="7q21.3-q22.1"
1..147
/gene="COL1A2"
/note="alpha-2 type I procollagen"
/codon_start=1
/db_xref="GDB:C00-119-062"
/protein_id="AA51887.1"
/db_xref="GI:179601"
/translation="KKTNEMGKTIIEYKTKNSALPFLDIAPLDIGADHFEFVDIG VCFK"

CDS

BASE COUNT 373 a 199 c 189 g 369 t
ORIGIN about 7.5 kb after <humc1a21>.

Query Match 57.2%; Score 61.8; DB 9; Length 1130;

Best Local Similarity 90.6%; Pred. No. 0.23;

Matches 77; Conservative 0; Mismatches 7; Indels 1; Gaps 1;

QY 6 ttgtatattgagatgtttaataaa-tgtgaaaaaaatgaaataagcatgtttggt 64

Db 918 TTGTATATGTGAGATCTTAATAATGTGAAAAAAGAAATGAATGAATGATGTTGGTT 977

QY 65 ttccaaagaaaaaataaaaaa 89

Db 978 TTCCAAAAGAACATATTGACTAAA 1002

RESULT 3

AF004877

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

AF004877 3682 bp DNA PRI 01-JAN-1998
Homo sapiens pro-alpha 2(I) collagen (COL1A2) gene, complete cds.
AF004877 AF004877.1 GI:2735714
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (sites)
Myers,J.C., Dickson,L.A., de Wet,W.J., Bernard,M.P., Chu,M.L., Di Liberto,M., Pepe,G., Sangiorgi,F.O. and Ramirez,F.
Analysis of the 3' end of the human pro-alpha 2(I) collagen gene.
utilization of multiple polyadenylation sites in cultured fibroblasts
J. Biol. Chem. 258 (16), 10128-10135 (1983)
MEDLINE 83290853
REFERENCE 2 (sites)

protein_bind 2487..2500
/gene="COL1A2"
/note="putative"
/citation=[7]
/bound_moiety="NF1"
2502..5143
/gene="COL1A2"
/citation=[11]
/citation=[7]
/number=1
3380..3386
/note="putative; bottom strand"
/bound_moiety="API"
3407..3413
/gene="COL1A2"
/note="putative"
/citation=[7]
/bound_moiety="API"
3716..3747
/citation=[7]
/rpl_type=tandem
/rpl_unit=gt
5144..5154
/gene="COL1A2"
/number=2
5155..5745
/gene="COL1A2"
/citation=[11]
/number=2
5746..5760
/gene="COL1A2"
/number=3
5761..6410
/gene="COL1A2"
/citation=[11]
/number=3
exon
intron
repeat_region
exon
intron
Query Match 57.2%; Score 61.8; DB 9; Length 38682;
Best Local Similarity 90.6%; Pred. No. 0.2;
Matches 77; Conservative 0; Mismatches 7; Indels 1; Gaps 1;
QY 6 ttgttatcttgagatggttaataa-tgtgaaaaaatgaataaagcatgttggct 64
|||||
Db 38471 TTGTATATGTGAGATGTTAAATGTGAAAAAATGAATAAGCATGTTGGTT 38530
|||||
QY 65 ttccaaagaaaaaataaaaaa 89
|||||
Db 38531 TTCCAAAGAACATATGACTMAAA 38555
|||||
RESULT 4
AC002528 141120 bp DNA PRI 04-FEB-2000
LOCUS AC002528 Human BAC clone GSI-542D18 from 7q31-q32, complete sequence.
DEFINITION AC002528
AC002528.1 GI:2388554
VERSION AC002528.1 GI:2388554
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 141120)
AUTHORS Kalicki,J., Wamsley,P. and Gibson,A.
TITLE The sequence of H. sapiens BAC clone GSI-542D18
JOURNAL Unpublished
2 (bases 1 to 141120)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (10-SEP-1997) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
3 (bases 1 to 141120)
AUTHORS Waterston,R.
TITLE Direct Submission

JOURNAL Submitted (03-FEB-2000) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
4 (bases 1 to 141120)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (04-FEB-2000) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: saplens@watson.wustl.edu
----- Summary Statistics
Center project name: H.GS542D18

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.
This sequence was finished as follows unless otherwise noted:
all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.
MAPPING INFORMATION:
The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see
http://www.nhgri.nih.gov/DIR/GRB/CHR7 or send
mailto:egreen@nhgri.nih.gov
SOURCE INFORMATION:
This clone is from the first BAC library from Genome Systems, Inc. (<http://www.genomesystems.com>).
Cell line: Lymphoblastoid
Haplotypes: two
VECTOR: pBelOBAC
Selection: chloramphenicol
NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is GSI-56H18, 200 bp overlap. The actual start of this clone is at base position 1 of GSI-542D18; actual end is at 141120 of GSI-542D18. This clone is part of an unanchored island, orientation is unknown.
This clone contains STS's SMS53299 (NTD:91113722) and SMS5558 (NTD:9454686).
FEATURES
source
Location/Qualifiers
1..141120
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="7"
/clone="GSI-542D18"
/clone_lib="GSBAC1"
/map="7q31-q32"
<680..24823
/gene="COL1A2"
join(<680..733,2277..2321,2613..2666,2766..2810,3200..3253,3752..3850,3994..4038,4153..4251,4672..4725,4850..4957,5319..5372,5486..5584,6498..6551,7014..7112,7513..7566,8120..8173,8324..8377,8652..8705,9656..9700,10835..10933,12154..12261,12929..12982,13928..13981,14663..14716,14821..14874,14971..15078,15439..15492,16329..16382,17387..17548,18766..18873,19547..19654,20040..20093,20180..20287,20428..20481,20853..20960,21438..21491,21618..21725,22057..22315,22723..22907,

TITLE
JOURNAL
COMMENT

Macdonald, P., Marquis, N., McEwan, P., McGurk, A., McKernan, K.,
McPheeters, R., Meldrum, J., Meneses, L., Morrow, J., Naylor, J.,
Norman, C.H., O'Connor, T., O'Donnell, P., Oliver, T.M., Peterson, K.,
Pierre, N., Pisan, C., Pollara, V., Raymond, C., Riley, R., Rotman, D.,
Roy, A., Santos, R., Severy, P., Spencer, B., Stange-Thomann, N.,
Stojanovic, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J.,
Tirrell, A., Vassiliev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W.J.,
Zimmer, A., and Zody, M.

Submitted (06-FEB-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Sep 20, 2000 this sequence version replaced g1:9108379.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu

Project Information
Center project name: L1396
Center clone name: 65_B_13

Summary Statistics
Sequencing vector: M13; M7815; 100% of reads
Assembly: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 179493 bases at least Q40
Consensus quality: 183982 bases at least Q30
Consensus quality: 185689 bases at least Q20
Insert size: 200000; sum-of-ctrls
Insert size: 186681; sum-of-ctrls
Quality coverage: 4.5 in Q20 bases: agarose-fp
Quality coverage: 4.8 in Q20 bases: sum-of-ctrls

NOTE: This is a 'working draft' sequence. It currently
consists of 22 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.

```

1      3528: contig of 3528 bp in length
*      3529 3628: gap of 100 bp
*      3629 5349: contig of 1721 bp in length
*      5350 5449: gap of 100 bp
*      5450 7196: contig of 1747 bp in length
*      7197 7296: gap of 100 bp
*      7297 9601: contig of 2305 bp in length
*      9602 9701: gap of 100 bp
*      9702 11931: contig of 2230 bp in length
*      11932 12031: gap of 100 bp
*      12032 14313: contig of 2282 bp in length
*      14314 14413: gap of 100 bp
*      14414 16778: contig of 2365 bp in length
*      16779 16878: gap of 100 bp
*      16879 19917: contig of 3039 bp in length
*      19918 20017: gap of 100 bp
*      20018 22727: contig of 2710 bp in length
*      22728 22827: gap of 100 bp
*      22828 28882: contig of 6055 bp in length
*      28883 28982: gap of 100 bp
*      28983 34336: contig of 5354 bp in length
*      34337 34436: gap of 100 bp
*      34437 40444: contig of 6008 bp in length
*      40445 40544: gap of 100 bp
*      40545 48669: contig of 8125 bp in length
*      48670 48769: gap of 100 bp
*      48770 57302: contig of 8533 bp in length
*      57303 57402: gap of 100 bp
*      57403 65250: contig of 7848 bp in length
*      65251 65350: gap of 100 bp

```

```

*      65351 89215: contig of 2365 bp in length
*      89216 89315: gap of 100 bp
*      89316 100430: contig of 11115 bp in length
*      100431 100530: gap of 100 bp
*      100531 115472: contig of 14942 bp in length
*      115473 115572: gap of 100 bp
*      115573 132563: contig of 16991 bp in length
*      132564 132663: gap of 100 bp
*      132664 152008: contig of 19345 bp in length
*      152009 152108: gap of 100 bp
*      152109 178451: contig of 26343 bp in length
*      178452 178551: gap of 100 bp
*      178552 188781: contig of 10230 bp in length.
*      Location/Qualifiers
1. 188781
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="15"
/map="15"
/clone="RP11-65B13"
/clone_lib="RP11 Human Male BAC"
1. 3528
/feature="assembly-fragment"
vector_side="left"
vector_end="SP6"
/feature="assembly-fragment"
3629..5349
/feature="assembly-fragment"
5450..7196
/feature="assembly-fragment"
7297..9601
/feature="assembly-fragment"
9702..11931
/feature="assembly-fragment"
12032..14313
/feature="assembly-fragment"
14414..16778
/feature="assembly-fragment"
16879..19917
/feature="assembly-fragment"
20018..22727
/feature="assembly-fragment"
22828..28882
/feature="assembly-fragment"
28983..34336
/feature="assembly-fragment"
34437..40444
/feature="assembly-fragment"
40545..48669
/feature="assembly-fragment"
48770..57302
/feature="assembly-fragment"
57403..65250
/feature="assembly-fragment"
65351..89215
/feature="assembly-fragment"
89316..100430
/feature="assembly-fragment"
100531..115472
/feature="assembly-fragment"
115573..132563
/feature="assembly-fragment"
132664..152008
/feature="assembly-fragment"
152109..178451
/feature="assembly-fragment"
178552..188781
/feature="assembly-fragment"
vector_side="right"

```

```

BASE COUNT 56489 a 35268 c 32583 g 62340 t 2101 others
ORIGIN

```


Query Match 57.2% Score 61.8; DB 2; Length 188781;
 Best Local Similarity 90.6%; Pred. No. 0.19;
 Matches 77; Conservative 0; Mismatches 7; Indels 1; Gaps 1;

QY 6 ttgtatattgagatgttcaataa-tgtgaaaaaaatgaataaagcattgtgtt 64
 |||
 Db 82663 TTTGTATATGTGAGATGTTAAATAATTGTGAATAAATAAATGAATGAATGAATGTTGCTT 82722

OY 65 ttccaaagaaaaaataaaaaa 89
 |||
 Db 82723 TTCCAAAAGACATATTGAGTAA 82747

RESULT 6
 AC092511 189401 bp DNA HTG 13-JUL-2001
 LOCUS Papio cynocephalus anubis clone RP41-444H21, WORKING DRAFT
 DEFINITION AC092511.1 GI:14717324
 AC092511
 VERSION AC092511.1
 KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT.
 SOURCE Olive baboon.
 ORGANISM Papio cynocephalus anubis
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
 Cercopithecinae; Papio.
 1 (bases 1 to 189401)
 Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B., Blakesley, R.W.,
 Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S., Dietrich, N.L.,
 Grantham, S., Guan, X., Gupta, J., Ho, S.-L., Idol, J.R., Karlins, E.,
 Lee-Jin, S.-O., Legaspi, R., Lim, M., Maduro, O.L., Maduro, V.B.,
 Mastello, C., Mastrian, S.D., McCloskey, J.C., McDowell, J.,
 Pearson, R., Prasad, A., Shevchenko, Y., Snyder, B., Stantirpop, S.,
 Thomas, J.W., Thomas, P.J., Touchman, J.W., Tsugeon, C., Vogt, J.L.,
 Walker, M.A., Wehrby, K.D., Zhang, L.-H. and Green, E.D.
 NISC Comparative Sequencing Initiative
 Unpublished
 2 (bases 1 to 189401)
 Green, E.D.
 Direct Submission
 Submitted (13-JUL-2001) NIH Intramural Sequencing Center, 8717
 Government Circle, Gaithersburg, MD 20877, USA
 ----- Genome Center
 Center: NIH Intramural Sequencing Center
 Center code: NISC
 Web site: http://www.nisc.nih.gov
 Contact: nisc_mouse@nhgri.nih.gov
 ----- Project Information
 Center project name: amn
 Center clone name: 444H21
 ----- Summary Statistics
 Sequencing vector: plasmid; n/a; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990319
 Consensus quality: 189312 bases at least Q40
 Consensus quality: 189377 bases at least Q40
 Consensus quality: 189398 bases at least Q20
 Insert size: 15400; agarose-ff
 Insert size: 19800; pulse-field-gel
 Insert size: 189401; sum-of-ctnigs
 Quality coverage: 15.10x in Q20 bases; pulse-field-gel
 Quality coverage: 11.74x in Q20 bases; pulse-field-gel
 Quality coverage: 12.28x in Q20 bases; sum-of-ctnigs

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 1 contigs. Gaps between the contigs
 * are represented as runs of N. The order of the pieces
 * is believed to be correct as given, however the sizes
 * of the gaps between them are based on estimates that have
 * provided by the submitter.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.

FEATURES
 source
 1 189401: contig of 189401 bp in length.
 Location/Qualifiers
 1..189401
 /organism="Papio cynocephalus anubis"
 /db_xref="taxon:9535"
 /clone="RP41-444H21"
 /clone_1id="RP41"
 1..189401
 /note="assembly fragment"

BASE COUNT 60088 a 36311 c 35278 g 57724 t
 ORIGIN

Query Match 57.2% Score 61.8; DB 2; Length 189401;
 Best Local Similarity 90.6%; Pred. No. 0.19;
 Matches 77; Conservative 0; Mismatches 7; Indels 1; Gaps 1;

QY 6 ttgtatattgagatgttcaataa-tgtgaaaaaaatgaataaagcattgtgtt 64
 |||
 Db 96550 TTTGTATATGTGAGATGTTAAATAATTGTGAATAAATAAATGAATGAATGTTGCTT 96609

OY 65 ttccaaagaaaaaataaaaaa 89
 |||
 Db 96610 TTCCAAAAGACATATTGAGTAA 96634

RESULT 7
 G15893
 LOCUS human STS CHLC_U01613_J03464.p56092 clone UTR_01613_J03464.
 DEFINITION G15893
 ACCESSION G15893
 VERSION G15893.1 GI:1161782
 STS sequence: primer: sequence tagged site.
 human vector=pUCPI host=E.coli dut1ung+ (DH10B) Marker Selected
 genomic DNA prepared from XY individual of French nationality.
 SOURCE Homo sapiens
 ORGANISM Homo sapiens
 Eukaryota; Eukaryota; Metazoa; Chordata;
 Vertebrata; Gnathostomata; Osteichthyes; Sarcopterygii; Chonata;
 Tetrapoda; Amniota; Mammalia; Theria; Eutheria; Archonta; Primates;
 Catarrhini; Hominoidea; Homo.
 1 (bases 1 to 846)
 Murray, J., Sheffield, V., Weber, J.L., Duyk, G. and Buettow, K.H.
 Cooperative Human Linkage Center
 Unpublished (1995)
 Synonyms: UTR_01613_J03464, CHLC_UTR_01613_J03464.T36142
 Contact: Dr. Jeffrey C. Murray
 UofI
 The University of Iowa
 Department of Pediatrics, Iowa City, IA 52242, USA
 Tel: (319) 356-3508
 Fax: (319) 356-3347
 Email: jeff-murray@uiowa.edu
 Primer A: TCACATGTTAGTGCCTGA
 Primer B: TTTGGCAATATATCTGCAT
 STS size: 247
 PCR Profile:
 denature: 30 seconds at 94 degrees C
 annealing: 75 seconds at 55 degrees C
 extension: 15 seconds at 72 degrees C
 PCR cycles: 27
 extension: 6 minutes at 72 degrees C
 Protocol:
 Template: 30ng genomic DNA
 Primer: each 1.5 pmole
 dNTPs: each 200 uM
 Tag Polymerase: 0.3 units
 Total Vol: 10 uL
 Buffer:
 MgCl2: 1.5mM
 KCl: 50mM
 Tris: 10mM

exon /number-13
779..832
/number-14
833..877
exon /number-15
878..931
/number-16
932..1030
exon /number-17
1031..1075
/number-18
1076..1174
exon /number-19
1175..1228
/number-20
1229..1336
exon /number-21
1337..1390
/number-22
1391..1489
exon /number-23
1490..1543
/number-24
1544..1642
exon /number-25
1643..1696
/number-26
1697..1750
exon /number-27
1751..1804
/number-28
1805..1858
exon /number-29
1859..1903
/number-30
1904..2002
exon /number-31
2003..2110
/number-32
2111..2164
exon /number-33
2165..2218
/number-34
2219..2272
exon /number-35
2273..2326
/number-36
2327..2434
exon /number-37
2435..2488
/number-38
2489..2542
exon /number-39
2543..2704
/number-40
2705..2812
exon /number-41
2813..2920
/number-42
2921..2974
exon /number-43
2975..3082
/number-44
3083..3136
exon /number-45
3137..3244
/number-46
3245..3298
exon /number-47
3299..3406
/number-48
3407..3665
exon /number-49

misc_feature 3452..3496
/note="C_telopectide"
misc_feature 3497..4237
/note="C_propeptide"
exon 3666..3850
/number=50
exon 3851..4093
/number=51
misc_feature 3938..3946

Query Match 55.4%; Score 59.8; DB 9; Length 5086;
Best Local Similarity 96.0%; Pred.No. 0.44;
Matches 72; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

Oy 6 ttgtatattgtgagatgtttaataaa-tgtgaataaataaataaagcatgtttggtt 64
|||||
Db 5011 TTGTATATGTGAGATGTTAAATAATTGTGAAAAAATGAATAACATGTTTGTT 5070
|||||

Oy 65 ttccaaagaataaa 79
|||||

Db 5071 TTCCAAAAGACACTA 5085

RESULT 9

HUMCIA2 5416 bp mRNA PRI 31-OCT-1994
LOCUS Human collagen alpha 2 type I mRNA, complete cds, clone pHC012A1.
DEFINITION J03464 M18057 X02488
ACCESSION J03464.1 GI:179595
VERSION alpha-2 type I collagen; collagen.
KEYWORDS alpha-2 type I collagen; collagen.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 751)
Dickson,L.A., de Wet,W., Di Liberto,M., Well,D. and Ramirez,F.
Analysis of the promoter region and the N-propeptide domain of the
human pro alpha 2(I) collagen gene
Nucleic Acids Res. 13 (10), 3427-3438 (1985)
85242047
2 (bases 335 to 5416)
de Wet,W., Bernard,M., Benson-Chanda,V., Chu,M.L., Dickson,L.,
Well,D. and Ramirez,F.
Organization of the human pro-alpha 2(I) collagen gene
J. Biol. Chem. 262 (33), 16032-16036 (1987)
88058962
3 (sites)
Wirtz,M.K., Gianville,R.W., Steinmann,B., Rao,V.H. and
Hollister,D.W.
Ehlers-Danlos syndrome type VIIB. Deletion of 18 amino acids
comprising the N-telopectide region of a pro-alpha 2(I) chain
J. Biol. Chem. 262 (34), 16376-16385 (1987)
88059013
4 (sites)
Well,D., Bernard,M., Combates,N., Wirtz,M.K., Hollister,D.W.,
Steinmann,B. and Ramirez,F.
Identification of a mutation that causes exon skipping during
collagen pre-mRNA splicing in an Ehlers-Danlos syndrome variant
J. Biol. Chem. 263 (18), 8561-8564 (1988)
88243702
5 (sites)
Marini,J.C., Lewis,M.B., Wang,Q., Chen,K.C. and Orrison,B.M.
Serine for glycine substitutions in type I collagen in two cases of
type IV osteogenesis imperfecta
J. Biol. Chem. (1992) in press
[2] revises [1].
[3] sites: Ehlers-Danlos syndrome type VIIB deletion. [4] sites:
polymorphism site.
Draft entry and computer-readable sequence [2] kindly provided by
F.Ramirez 18-OCT-1987.
EMBL features not translated to GenBank features:
key from pot. CANT-box
PRM 249 253

Query Match	51.9%	Score 56	DB 2	Length 236699
Best Local Similarity	86.9%	Pred. No. 1.5		
Matches 73	Conservative 0	Mismatches 10	Indels 1	Gaps
OY	6	tttgatactgagatgttcaataaatgtcaaaaaatgaataagcatgttgctt	65	
Db	230253	TTTGATATGTCGACATGCTTTAAATAAAT-TGTGAAAAAATAAATAAGCATTTTGTGCTT	230311	
OY	66	tccaaaagaaaaaataaaaaaa 89		
Db	230312	TCCAAAAAGAAAAATATGTGCTAAAA	230335	
RESULT 11				
AF117707				
LOCUS	AF117707	1162 bp	mRNA	PLN
DEFINITION	Lycopersicon esculentum putative copper/zinc superoxide dismutase			
	copper chaparon precursor (CCS) mRNA, partial cds; nuclear gene			
	for chloroplast product.			
ACCESSION	AF117707			
VERSION	AF117707.2	GI:5759319		
KEYWORDS	.			
SOURCE	tomato.			
ORGANISM	Lycopersicon esculentum			
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;			
	Asteralide; euasterid I; Solanales; Solanaceae; Solanum;			
REFERENCE	Lycopersicon.			
AUTHORS	1 (bases 1 to 1162)			
TITLE	Nersissian,A.M. and Valentine,J.S.			
	Full-length cDNA encoding tomato Lys7, copper chaparon for Cu,Zn			
	superoxide dismutase			
JOURNAL	Unpublished			
REFERENCE	2 (bases 155 to 1162)			
AUTHORS	Nersissian,A.M. and Valentine,J.S.			
TITLE	Direct Submission			
JOURNAL	Submitted (04-JAN-1999) Chemistry & Biochemistry, University of			
	California Los Angeles, 405 Hilgard Ave., Los Angeles, CA 90095,			
	USA			

XX AL357036.8
SV 27-MAY-2000 (Rel. 63, Created)
XX 23-JAN-2001 (Rel. 66, Last updated, Version 8)
DT
XX Human DNA sequence *** SEQUENCING IN PROGRESS *** from clone RP11-565J7
XX HTG: HTGS_DRAFT; HTGS_PHASE1.
XX
XX Homo sapiens (human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia;
XX Eutheria; Primates; Catarrhini; Hominidae; Homo.
XX
XX [1]
RA Plumb B.;
RT Submitted (21-JAN-2001) to the EMBL/GenBank/DBJ databases.
RL Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
RL humquery@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk
XX
XX ----- Genome Center
CC Center: Sanger Centre
CC Center code: SC
CC Web site: http://www.sanger.ac.uk
CC Contact: humquery@sanger.ac.uk
CC ----- Project Information
CC Center project name: ba56537
CC ----- Summary Statistics
CC Assembly program: XGAP4; Version 4.5
CC Sequencing vector: plasmid; L08752; 100% of reads
CC Chemistry: Dye-terminator Big Dye; 100% of reads
CC Consensus quality: 141827 bases at least Q40
CC Consensus quality: 146636 bases at least Q30
CC Insert size: 149144; sum-of-contigs
CC Insert size: 168730; 3.3% error; agarose-fp
CC Quality coverage: 4.18x in Q20 bases; sum-of-contigs
CC Quality coverage: 4.07x in Q20 bases; agarose-fp
CC
CC -----
CC * NOTE: This is a 'working draft' sequence. It currently
CC consists of 18 contigs. The true order of the pieces is
CC * not known and their order in this sequence record is
CC arbitrary. Where the contigs adjacent to the vector can
CC * be identified, they are labelled with 'clone_end' in the
CC feature table. Some order and orientation information
CC * can tentatively be deduced from paired sequencing reads
CC which have been identified to span the gap between two
CC * contigs. These are labelled as part of the same
CC * 'fragment_chain', and the order and relative orientation
CC * of the pieces within a fragment_chain is reflected in
CC this file. Gaps between the contigs are represented as
CC * runs of N, but the exact sizes of the gaps are unknown.
CC * This record will be updated with the finished sequence as
CC * soon as it is available and the accession number will be
CC * preserved.
CC
CC 1 13329 contig of 13329 bp in length
CC 13330 24463 contig of 11134 bp in length; fragment_chain 1
CC 24564 35124 contig of 10561 bp in length; fragment_chain 1
CC 35225 43383 contig of 8159 bp in length; fragment_chain 1
CC 43484 50759 contig of 7276 bp in length; fragment_chain 1
CC 50860 60970 contig of 10111 bp in length; fragment_chain 1
CC 61071 66063 contig of 4993 bp in length; fragment_chain 2
CC 66164 68816 contig of 3653 bp in length; fragment_chain 2
CC 6917 74097 contig of 4181 bp in length; fragment_chain 2
CC 74198 77759 contig of 3562 bp in length; fragment_chain 3
CC 77860 86114 contig of 8255 bp in length; fragment_chain 3
CC 86215 98174 contig of 11960 bp in length; fragment_chain 3
CC 98275 105363 contig of 7089 bp in length; fragment_chain 3
CC 105466 113505 contig of 9842 bp in length; fragment_chain 4
CC 115406 120735 contig of 5330 bp in length; fragment_chain 4
CC 120836 133308 contig of 12373 bp in length; fragment_chain 4
CC 133309 148679 contig of 15371 bp in length

CC * 148780 150844 contig of 2065 bp in length
XX
FH Key Location/Qualifiers
FT source 1..150844
FT /chromosome="1"
FT /db_xref="taxon:9606"
FT /organism="Homo sapiens"
FT /clone="RP11-565J7"
FT /clone_11b="RPCI-11.2"
FT 1..13329
FT /note="assembly_fragment:01838"
FT /note="clone_end:SP6"
FT /note="vector_side:left"
FT 13330..24463
FT /note="assembly_fragment:01484"
FT /note="fragment_chain:1"
FT 24564..35124
FT /note="assembly_fragment:01776"
FT /note="fragment_chain:1"
FT 35225..43383
FT /note="assembly_fragment:01297"
FT /note="fragment_chain:1"
FT 43484..50759
FT /note="assembly_fragment:01491"
FT /note="fragment_chain:1"
FT 50860..60970
FT /note="assembly_fragment:01931"
FT /note="fragment_chain:1"
FT 61071..66063
FT /note="assembly_fragment:00043"
FT /note="fragment_chain:2"
FT 66164..6917
FT /note="assembly_fragment:00370"
FT /note="fragment_chain:2"
FT 6917..74097
FT /note="assembly_fragment:01845"
FT /note="fragment_chain:2"
FT 74198..77759
FT /note="assembly_fragment:01000"
FT /note="fragment_chain:3"
FT 77860..86114
FT /note="assembly_fragment:01842"
FT /note="fragment_chain:3"
FT 86215..98174
FT /note="assembly_fragment:01350"
FT /note="fragment_chain:3"
FT 98275..105363
FT /note="assembly_fragment:01525"
FT /note="fragment_chain:4"
FT 105464..115305
FT /note="assembly_fragment:00635"
FT /note="fragment_chain:4"
FT 115406..120735
FT /note="assembly_fragment:01073"
FT /note="fragment_chain:4"
FT 120836..133208
FT /note="assembly_fragment:00972"
FT 133309..148679
FT /note="assembly_fragment:01642"
FT 148780..150844
FT /note="assembly_fragment:00704"
FT /note="clone_end:T7"
FT /note="vector_side:right"
XX
SQ Sequence 150844 BP; 44269 A; 29934 C; 30171 G; 44748 T; 1722 other;

Query Match 49.6%; Score 53.6; DB 33; Length 150844;
Best Local Similarity 77.4%; Pred. No. 3.5;
Matches 65; Conservative 0; Mismatches 19; Indels 0; Gaps 0;
Oy 20 atgtttaataaagtgaataaataaagaatgaatgattgtttcccaaaagaaaaa 79

```

Db 79337 ATGCATTAATAGTGTTAAAAAATAAGACCTCTGGGCTAAAAA 79396
QY 80 aaaaaaaaaaaaaaaaaa 103
Db 79397 AAAAAAAAAAAAAAAAAA 79420

```

```

RESULT 15
AC092814/c 169224 bp DNA HTG 28-JUL-2001
LOCUS Homo sapiens chromosome 1 clone RP11-565J7, WORKING DRAFT SEQUENCE,
DEFINITION 3 unordered pieces.
ACCESSION AC092814 AL357036
VERSION AC092814.1 GI:15027767
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP; HTGS_ACTIVEPIN.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 169224)
Kaul,R.K., Olson,M.V., Raymond,C., Clendinning,J., Ivey,R.G. and
Haugen,E.D.
REFERENCE Unpublished
AUTHORS Direct Submission
2 (bases 1 to 169224)
Kaul,R.K., Olson,M.V., Raymond,C., Clendinning,J., Ivey,R.G. and
Haugen,E.D.
TITLE Direct Submission
JOURNAL Unpublished
AUTHORS Submitted (28-JUL-2001) Genome Center, University of Washington,
Box 352145, Seattle, WA 98195, USA
COMMENT On Jul 28, 2001 this sequence version replaced gi:12539675.
----- Genome Center
Center: University of Washington Genome Center
Center Code: UMG
Web site: http://www.genome.washington.edu
Contact: uwgchgs@u.washington.edu
Drafting Center: SC
----- Project Information
Center project name: chr-1
Center clone name: RP11-565J7 (sc0180)
----- Summary Statistics
Sequencing vector: plasmid; 46% of reads
Sequencing vector: plasmid; 108752; 54% of reads
Chemistry: Dye-terminator Big Dye; 89% of reads
Chemistry: Dye-terminator Big Dye; 11% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 167317 bases at least Q40
Consensus quality: 168242 bases at least Q30
Consensus quality: 168768 bases at least Q20
Insert size: 169024; sum-of-ctrls
Quality coverage: 8.3x in Q20 bases; sum-of-ctrls

```

```

* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

```

1 16209: contig of 16209 bp in length
* 16210 16309: gap of unknown length
* 16310 59803: contig of 43494 bp in length
* 59804 59903: gap of unknown length
* 59904 169224: contig of 109321 bp in length.
Location/Qualifiers
1. 169224
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP11-565J7"
/clone_1tb="RPC1 human BAC library 11"
FEATURES
SOURCE

```

```

misc_feature 1..16209
/note="assembly_name:Contig15"
misc_feature 16310..59803
/note="assembly_name:Contig16"
misc_feature 59904..169224
/note="assembly_name:Contig17"
BASE COUNT 48926 a 33490 c 34553 g 51952 t 303 others
ORIGIN

```

```

Query Match 49.6%; Score 53.6; DB 2: Length 169224;
Best Local Similarity 77.4%; Pred. No. 3.5;
Matches 65; Conservative 0; Mismatches 19; Indels 0; Gaps 0;
QY 20 atgtttaataatgtaaaaaataaataaagcagttgtttccaaagaaaaa 79
Db 53306 ATGCATTAATAGTGTTAAAAAATAAGACCTCTGGGCTAAAAA 53247
QY 80 aaaaaaaaaaaaaaaaaa 103
Db 53246 AAAAAAAAAAAAAAAAAA 53223

```

```

Search completed: January 17, 2002, 08:48:53
Job time: 10221 sec

```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 17, 2002, 06:42:50 ; Search time 2644.98 Seconds
(without alignments)
426.564 Million cell updates/sec

Title: US-09-823-101-7

Perfect score: 105

Sequence: 1 tgcagtggtgcagacatcc.....ggccgcctctctgaagctt 105

Scoring table:

IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*
1: em_estfun:*
2: em_esthum:*
3: em_estin:*
4: em_estom:*
5: em_estpl:*
6: em_estba:*
7: em_estro:*
8: em_estov:*
9: em_htc:*
10: gb_estl:*
11: gb_est2:*
12: gb_htc:*
13: gb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_inv:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_rnd:*
20: em_gss_vit:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	31	29.5	545	13	AZ446318 1M0242M08
C 2	30.6	29.1	551	13	AQ499366 HS_5199_B
C 3	30.6	29.1	601	13	AZ415460 1M0190E03
C 4	30.2	28.8	260	11	D81803 HUM409B05B
C 5	30.2	28.8	331	11	H91634 yv03f12.s1
C 6	30.2	28.8	359	10	AM006058 w281d05.x
C 7	30.2	28.8	361	10	AA252861 zt76c01.s
C 8	30.2	28.8	386	10	BE223043 hu46d12.x
C 9	30.2	28.8	394	10	AA460473 zx61e07.r
C 10	30.2	28.8	402	10	AA130279 z129c07.r
C 11	30.2	28.8	404	11	R76040 y160c01.r1
C 12	30.2	28.8	416	10	AV725870 AV725870

13	30.2	28.8	418	10	A1076519
C 14	30.2	28.8	422	11	N90308
C 15	30.2	28.8	422	11	N94771
C 16	30.2	28.8	424	10	A1033293
C 17	30.2	28.8	429	10	A1200105
C 18	30.2	28.8	432	11	w73949
C 19	30.2	28.8	443	10	AA017343
C 20	30.2	28.8	447	10	A1275117
C 21	30.2	28.8	448	10	A1131149
C 22	30.2	28.8	480	10	AA461041
C 23	30.2	28.8	486	10	A1955225
C 24	30.2	28.8	498	10	AA782083
C 25	30.2	28.8	514	10	A1382006
C 26	30.2	28.8	537	10	A1752305
C 27	30.2	28.8	546	10	AM007337
C 28	30.2	28.8	561	10	BE002118
C 29	30.2	28.8	582	10	A1798790
C 30	30.2	28.8	625	10	AV658702
C 31	30.2	28.8	679	10	AV731307
C 32	30.2	28.8	699	10	AV731039
C 33	30.2	28.8	721	10	AV731332
C 34	30.2	28.8	938	10	AL573092
C 35	30	28.6	755	13	CNS01P9L
C 36	29.6	28.2	933	13	CNS03WSJ
C 37	29.4	28.0	508	10	AM634290
C 38	29.4	28.0	534	10	AM460975
C 39	29.2	27.8	559	10	AA019909
C 40	29	27.6	471	13	AO735408
C 41	28.8	27.4	294	12	AK019058
C 42	28.8	27.4	456	12	AK008062
C 43	28.8	27.4	456	12	AK008159
C 44	28.8	27.4	456	12	AK008313
C 45	28.8	27.4	456	12	AK008420

ALIGNMENTS

RESULT 1
AZ446318/c
LOCUS
DEFINITION 1M0242M08R Mouse 10kb plasmid U0GC1M library Mus musculus genomic
clone U0GC1M0242M08 R, DNA sequence.

ACCESSION
AZ446318
VERSION
KEYWORDS
SOURCE
ORGANISM

house mouse.
Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
AUTHORS
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamll,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,
M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
and Wright,D., Weiss,R.

TITLE
Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts

JOURNAL
COMMENT
Unpublished (2000)
Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah

Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunne@genetics.utah.edu

Insert length: 10000 Std Error: 0.00
Plate: 0242 row: M column: 08

Seq primer: CACACGAGAAACACCTATGACC
Class: plasmid ends
High quality sequence stop: 545.

FEATURES
SOURCE
Location/Qualifiers
1..545

BASE COUNT
ORIGIN

163 a 150 c 120 g 112 t

```

/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="U0GC1M0242M08"
/clone_1ib="Mouse 10kb plasmid U0GC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PMD42nv, Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adaptored DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of PMD42 (g147321141gb1AF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adaptored mouse DNA was annealed to
adaptored vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."
```

Query Match	29.5%	Score 31	DB 13	Length 545
Best Local Similarity	62.0%	Pred. No. 15		
Matches 49	Conservative	0	Mismatches 30	Indels 0
				Gaps
Oy	5	ggtgtgcagacatcctccagttgtgtgtacagatgacgcgcagcagtgtgtatgtg	64	
Db	480	gntttgcacagacacctctgcagcttctctgtatgntcctgtracagacgctgttttg	422	
Oy	65	ctgcctcttcacatgagat	83	
Db	420	ctgcctgtttcacacttgagg	402	

RESULT	2
LOCUS	A0499366/c
DEFINITION	A0499366 551 bp DNA GSS 28-Apr-1999 HS_5199_B2_D02_SF6E_RPCT-11 Human Male BAC Library Homo sapiens genomic clone Plate=775 Col=4 Row=W, DNA sequence.
ACCESSION	A0499366
VERSION	A0499366.1 GI:4699478
KEYWORDS	GSS.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. I (bases 1 to 551)
AUTHORS	Mahairas,G.G., Wallace,J.C., Smith,R., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J.J., Young,J.J., Zhao,S., Adams,M.D. and Hood,L.
TITLE	Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
JOURNAL	Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE	99380589
COMMENT	Contact: Mahairas GG, Wallace JC, Hood L

Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPEC-11. For BAC library availability, please contact Pierer de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from

```

BACpac Resources (http://bacpac.med.buffalo.edu/ordering\_bac.htm)
or from Researc h Genetics (http://www.rhsc.washington.edu)
Plate: 775   row:  H   column:  4
Seq primer:  SP6
Class:  BAC ends
High quality sequence stop: 551.
Location/Qualifiers
1..551

```

BASE COUNT	143 a	167 c	115 g	124 t	2 others
ORIGIN					

Query Match	29.1%	Score 30.6	DB 13	Length 551
Best Local Similarity	58.1%	Pred. No. 20		
Matches 54	Conservative 0	Mismatches 39	Indels 0	Gaps 0

OY 13 agacatctgcgaatttcgtagagataaacagacgaagtgtgatltgcggcctt 72
| | | | | | | | | | | | | | | | | | | | |
Db 217 ACACCTCCTGTAAAGTCATGCGTTTCCATGGGGCAACGATGTTCTGCCCTTGCTGCTCA 158

OY 73 tgcactagatcacaggcgccctctctcaagctt 105
| | | | | | | | | | | | | | | | | | | | |
Db 157 TGTTGAGGCTCATTTCCAAGTGTTATCAAGCTT 125

RESULT	3
AZ415460/c	
LOCUS	AZ415460
DEFINITION	601 bp DNA GSS 03-OCT-2000
	IM0190E03 Mouse 10kb plasmid UUC1M library Mus musculus genomic
	clone UUC1M0190E03 F, DNA sequence.

VERSION	A2415460.1	GI:10539473
KEYWORDS	GSS.	
SOURCE	house mouse,	
ORGANISM	Mus musculus	

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
Euarctota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 601)	Dunn, D., Aoyagi, A., Barber, M., Baacorn, T., Duval, B., Hamil, C., Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A. and Wright, D., Weiss, R.	Mouse whole genome scaffolding with paired end reads from 10Kb plasmid inserts	unpublished (2000)	Contact: Robert B. Weiss

Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
Tel.: 801 585 5606
Fax: 801 585 7177
Email: ddunne@genetics.utah.edu
Insert length: 10000 Std Error: 0.00
Plate: 0190 row: E column: 03
Seq primer: CGTTGTAAACGACGCGCAGT
Class: plasmid ends
High quality sequence stop: 601.

```

FEATURES
source      Location/Qualifiers
1. .601     /organism="Mus musculus"
            /strain="C57BL/6J"

```

/db.xref="taxon:10090"
/clone="UUGC1M0190E03"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, Tl-resistant, F-"
/muscle="Vector: PMD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adapted DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of pMDA2 (gII473211419b1AFI29072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adapted mouse DNA was annealed to
adapted vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."

BASE COUNT 239 a 204 c 76 g 82 t

ORIGIN

Query Match 29.1%; Score 30.6; DB 13; Length 601;
Best Local Similarity 60.0%; Pred. No. 20;

Matches 51: Conservative 0; Mismatches 34; Indels 0; Gaps 0;

Oy 17 attctgcacgttcttgtagcatgatcagcagcagtgtgttatgtctgtctttga 76
 ||| | | | | | | | | | | | | | | | | | | | | | | | |
Db 464 ATCCCCCTATGTGTTGTAGTGTATGATGCCCCCTGTGTATGTGTATGTCTGA 405
 ||| | | | | | | | | | | | | | | | | | | | | | | | |

Oy 77 ctgagatcaaggccgcccttccttaa 101
 ||| | | | | | | | | | | | | | | | | | | | | | | | |
Db 404 GTATGATGATCCCTCTGTGTGTGA 380

RESULT 4

LOCUS DB1803/c

DEFINITION HMW409B05B Human fetal brain (TFujiwara) Homo sapiens EST

ACCESSION GEN-409B05 5', mRNA sequence.

VERSION DB1803

CROSSREF DB1803.1 GI:1180434

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheraia; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 260)
Fujiwara,T., Hirano,H., Katagiri,T., Kawai,A., Kuwa,Y., Nagata,M.,
Okuno,S., Ozaki,K., Shimizu,F., Shinada,Y., Shinozumiya,H., Takachi,
A., Takeda,S., Watanabe,T., Takahashi,E., Hirai,Y., Maeawa,H.,
Shin,S. and Nakamura,Y.
Fujiwara et al. (1995)
Unpublished (1995)
Contact: Tsutomu Fujiwara
Otsuka GEN Research Institute
Otsuka Pharmaceutical Co.,Ltd
Tel: 0886-65-2888
Fax: 0886-37-1035.

TITLE JOURNAL

COMMENT Unpublished (1995)

FEATURES

source Location/Qualifiers

 1..260

 /organism="Homo sapiens"

 /db_xref="taxon:9606"

 /clone="GEN-409B05"

 /clone_lib="Human fetal brain (TFujiwara)"

 /dev_stage="fetal"

	BASE COUNT	85 a	/note="Organ: brain"	53 c	52 g	70 t	
	ORIGIN						
	Query Match	28.8%	Score 30.2;	DB 11:	Length 260;		
	Best Local Similarity	69.5%;	Pred. No. 21;				
	Matches 41;	Conservative 0;	Mismatches 18;	Indels 0;	Gaps 0;		
Oy	15 acatctgcacattctgtgtacagatgacgaacgatgtgtatgtcgccttt 73 						
Db	84 ACTCTCTGGAATTCTCTCTCATACATCTGCACAAGACTGGGTGGTGTTGTCCTT 26						
	RESULT 5	H91634	331 bp	mRNA	EST	29-NOV-1995	
	LOCUS H91634	yv03ff12.s1 Soares fetal liver spleen INFLS Homo sapiens CDNA clone					
	DEFINITION IMAGE:241679 3', mRNA sequence.						
	ACCESSION H91634						
	VERSION H91634.1	GI:1087212					
	KEYWORDS EST.						
	SOURCE human.						
	ORGANISM Homo sapiens						
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiide; Homo. 1 (bases 1 to 331) Hallier,L., Clark,N., Duboue,T., Elliston,K., Hawkins,M., Holman M., Hulthan,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevasakis,E., Waterston R., Williamson,A., Wohlmann,P. and Wilson,R. The Mashu-Merck EST Project Contact: Wilson RK Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: est@watson.wustl.edu High quality sequence stops: 291 Source: IMAGE Consortium, LNLN. This clone is available royalty-free through LNLN ; contact the IMAGE Consortium (InfoImage.lnl.n.gov) for further Information. Seq primer: Promega -21mi13 High quality sequence stop: 291. Location/Qualifiers 1. 331 /organism="Homo sapiens" /db_xref="GDB:3790812" /db_xref="taxon:9606" /clone="IMAGE:241679" /sex="male" /dev_stage="20 week-post conception fetus" /lab_host="DH10B (ampicillin resistant)" /note="Organ: Liver and Spleen; Vector: pTV73D (Pharmacia) with a modified polylinker; Site.1: Pac I; Site.2: Eco RI; 1st strand cDNA was primed with a Pac I - oligo(dT) primer [5', AACTCGAAGAATTAATTAAGATCTTTTCTTTTCTTTTCTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac I and cloned into the Pac I and Eco RI sites of the modified pTV73 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fátima Bonaldo."						
	FEATURES						
	SOURCE						
	BASE COUNT	92 a	64 c	63 g	106 t	6 others	
	ORIGIN						
	Query Match	28.8%	Score 30.2;	DB 11:	Length 331;		
	Best Local Similarity	69.5%;	Pred. No. 23;				
	Matches 41;	Conservative 0;	Mismatches 18;	Indels 0;	Gaps 0;		
Oy	15 acatctgcacattctgtgtacagatgacgaacgatgtgtatgtcgccttt 73						

Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL, send email to: info@image.llnl.gov

Seq primer: -40UP from GIBCO
High quality sequence stop: 344.

Location/Qualifiers

1. .386

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:317311"

/clone_lib="NCI-CGAP_Brn41"

/issue_type="oligodendroglioma"

/lab_host="DH10B (phage-resistant)"

/note="Organ: brain; Vector: pT73D-Pac (Pharmacia) with a modified polylinker. Site.1: Not I - oligo(dT) primer [5'

strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTCACATCTGAAGTGGAGCGCGCCATCATCTTTTATTTTATTTT 3'];

double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library

constructed and normalized by Bento Soares and M.Fatima

Bonaldo."

Bonaldo."

Bonaldo."

Bonaldo."

Bonaldo."

Bonaldo."

Bonaldo."

Bonaldo."

Bonaldo."

Bonaldo."

Bonaldo."

Bonaldo."

Bonaldo."

Bonaldo."

Bonaldo."

Bonaldo."

Bonaldo."

Bonaldo."

Bonaldo."

Bonaldo."

Bonaldo."

Bonaldo."

Bonaldo."

Bonaldo."

Bonaldo."

Bonaldo."

Bonaldo."

Bonaldo."

Bonaldo."

Bonaldo."

Bonaldo."

Bonaldo."

Bonaldo."

Bonaldo."

Bonaldo."

Bonaldo."

Bonaldo."

Bonaldo."

Bonaldo."

Bonaldo."

Bonaldo."

Bonaldo."

Bonaldo."

Bonaldo."

Bonaldo."

Bonaldo."

Bonaldo."

Bonaldo."

Bonaldo."

Bonaldo."

Bonaldo."

Bonaldo."

Bonaldo."

Bonaldo."

Bonaldo."

Bonaldo."

Bonaldo."

polylinker. Site.1: Not I; Site.2: Eco RI; 1st strand cDNA was prepared from mRNA obtained from pooled 8-9 week (total) fetus material with a Not I - oligo(dT) primer [5' TGTTCACATCTGAAGTGGAGCGCGCCATCATCTTTTATTTTATTTT 3'].

Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library

went through one round of normalization, and was

constructed by Bento Soares and M. Fatima Bonaldo."

constructed by Bento Soares and M. Fatima Bonaldo."

constructed by Bento Soares and M. Fatima Bonaldo."

constructed by Bento Soares and M. Fatima Bonaldo."

constructed by Bento Soares and M. Fatima Bonaldo."

constructed by Bento Soares and M. Fatima Bonaldo."

constructed by Bento Soares and M. Fatima Bonaldo."

constructed by Bento Soares and M. Fatima Bonaldo."

constructed by Bento Soares and M. Fatima Bonaldo."

constructed by Bento Soares and M. Fatima Bonaldo."

constructed by Bento Soares and M. Fatima Bonaldo."

constructed by Bento Soares and M. Fatima Bonaldo."

constructed by Bento Soares and M. Fatima Bonaldo."

constructed by Bento Soares and M. Fatima Bonaldo."

constructed by Bento Soares and M. Fatima Bonaldo."

constructed by Bento Soares and M. Fatima Bonaldo."

constructed by Bento Soares and M. Fatima Bonaldo."

constructed by Bento Soares and M. Fatima Bonaldo."

constructed by Bento Soares and M. Fatima Bonaldo."

constructed by Bento Soares and M. Fatima Bonaldo."

constructed by Bento Soares and M. Fatima Bonaldo."

constructed by Bento Soares and M. Fatima Bonaldo."

constructed by Bento Soares and M. Fatima Bonaldo."

constructed by Bento Soares and M. Fatima Bonaldo."

constructed by Bento Soares and M. Fatima Bonaldo."

constructed by Bento Soares and M. Fatima Bonaldo."

constructed by Bento Soares and M. Fatima Bonaldo."

constructed by Bento Soares and M. Fatima Bonaldo."

constructed by Bento Soares and M. Fatima Bonaldo."

constructed by Bento Soares and M. Fatima Bonaldo."

constructed by Bento Soares and M. Fatima Bonaldo."

constructed by Bento Soares and M. Fatima Bonaldo."

constructed by Bento Soares and M. Fatima Bonaldo."

constructed by Bento Soares and M. Fatima Bonaldo."

constructed by Bento Soares and M. Fatima Bonaldo."

constructed by Bento Soares and M. Fatima Bonaldo."

constructed by Bento Soares and M. Fatima Bonaldo."

constructed by Bento Soares and M. Fatima Bonaldo."

constructed by Bento Soares and M. Fatima Bonaldo."

constructed by Bento Soares and M. Fatima Bonaldo."

constructed by Bento Soares and M. Fatima Bonaldo."

constructed by Bento Soares and M. Fatima Bonaldo."

constructed by Bento Soares and M. Fatima Bonaldo."

constructed by Bento Soares and M. Fatima Bonaldo."

constructed by Bento Soares and M. Fatima Bonaldo."

constructed by Bento Soares and M. Fatima Bonaldo."

constructed by Bento Soares and M. Fatima Bonaldo."

constructed by Bento Soares and M. Fatima Bonaldo."

constructed by Bento Soares and M. Fatima Bonaldo."

constructed by Bento Soares and M. Fatima Bonaldo."

constructed by Bento Soares and M. Fatima Bonaldo."

constructed by Bento Soares and M. Fatima Bonaldo."

constructed by Bento Soares and M. Fatima Bonaldo."

constructed by Bento Soares and M. Fatima Bonaldo."

constructed by Bento Soares and M. Fatima Bonaldo."

constructed by Bento Soares and M. Fatima Bonaldo."

constructed by Bento Soares and M. Fatima Bonaldo."

constructed by Bento Soares and M. Fatima Bonaldo."

constructed by Bento Soares and M. Fatima Bonaldo."

constructed by Bento Soares and M. Fatima Bonaldo."

constructed by Bento Soares and M. Fatima Bonaldo."

constructed by Bento Soares and M. Fatima Bonaldo."

constructed by Bento Soares and M. Fatima Bonaldo."

constructed by Bento Soares and M. Fatima Bonaldo."

constructed by Bento Soares and M. Fatima Bonaldo."

constructed by Bento Soares and M. Fatima Bonaldo."

constructed by Bento Soares and M. Fatima Bonaldo."

constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 129 a 74 c 78 g 117 t 4 others
ORIGIN

Query Match 28.8%; Score 30.2; DB 10; Length 402;
Best Local Similarity 69.5%; Pred. No. 24;
Matches 41; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

Db 141 ACTTCCTGTGAGTTCCTCTGTGACATCTGCAAGACGCTGCTGGTGTGTCTTT 83

RESULT 11
R76040 404 bp mRNA EST 06-JUN-1995
LOCUS Y160C01.r1 Soares placenta Nb2HP Homo sapiens cDNA clone
DEFINITION IMAGE:143616 5', mRNA sequence.
ACCESSION R76040
VERSION R76040.1 GI:850722
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 404)
AUTHORS Haller, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevas, E., Waterston, R., Williamson, A., Wohlmann, P., and Wilson, R.
The WashU-Merck EST Project
Contact: Wilson RK
Unpublished (1995)
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Insert Size: 712
High quality sequence stops: 293
Source: IMAGE Consortium, LNL
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Insert Length: 712 Std Error: 0.00
Seq primer: M13RP1
High quality sequence stop: 293.

FEATURES
source Location/Qualifiers
1..404
/organism="Homo sapiens"
/db_xref="GDB:552797"
/db_xref="taxon:9606"
/clone="IMAGE:143616"
/clone_lib="Soares placenta Nb2HP"
/sex="female"
/dev_stage="Placenta obtained at birth (full term)"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: Placenta; Vector: pTZ19 (Pharmacia) with a
modified polylinker; Site: 1: Not I; Site 2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
AACTGACAGATTTTCGGCCGACGAGATTTTCTTTTCTTTT 3']
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pTZ19 vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT 128 a 72 c 82 g 120 t 2 others
ORIGIN

Query Match 28.8%; Score 30.2; DB 11; Length 404;
Best Local Similarity 69.5%; Pred. No. 24;
Matches 41; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

Db 140 ACTTCCTGTGAGTTCCTCTGTGACATCTGCAAGACGCTGCTGGTGTGTCTTT 82

RESULT 12
AV725870 416 bp mRNA EST 16-OCT-2000
LOCUS AV725870 HTC Homo sapiens cDNA clone HTCAMD10 5', mRNA sequence.
DEFINITION AV725870
ACCESSION AV725870
VERSION AV725870.1 GI:10831692
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 416)
AUTHORS Gu, Y., Peng, Y., Song, H., Huang, Q., Yang, Y., Gao, G., Xiao, H., Xu, X., Li, N., Qian, B., Liu, F., Qu, J., Gao, X., Cheng, Z., Xu, Z., Zeng, L., Xu, S., Gu, W., Tu, Y., Jia, J., Fu, G., Ren, S., Zhong, M., Lu, G., Hu, R., Chen, J., Chen, Z., and Han, Z.
Homo sapiens cDNA HTC clones
Unpublished (2000)
Contact: Zeguang Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
Tel: 86-21-50801919 (ex.45)
Fax: 86-21-50801922
Email: hanzg@chgc.sh.cn
This clone is available at CHGC in Shanghai.

FEATURES
source Location/Qualifiers
1..416
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="HTCAMD10"
/clone_lib="HTC"
/tissue_type="Hypothalamus"
/dev_stage="Adult"
/lab_host="SOLR"
/note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
XhoI"

BASE COUNT 131 a 78 c 77 g 128 t 2 others
ORIGIN

Query Match 28.8%; Score 30.2; DB 10; Length 416;
Best Local Similarity 69.5%; Pred. No. 24;
Matches 41; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

Db 68 ACTTCCTGTGAGTTCCTCTGTGACATCTGCAAGACGCTGCTGGTGTGTCTTT 10

RESULT 13
A1076519 418 bp mRNA EST 29-SEP-1998
LOCUS A1076519
DEFINITION OZ09104.x1 Soares fetal_liver_spleen_lnf15_S1 Homo sapiens cDNA
clone IMAGE:1674847 3', mRNA sequence.
ACCESSION A1076519
VERSION A1076519.1 GI:3405697
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 418)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 17, 2002, 08:48:57 ; Search time 2206.96 Seconds
(without alignments)
2900.327 Million cell updates/sec

Title: US-09-823-101-8

Perfect score: 388
Sequence: 1 acgcctggcgcacagcgtgagc.....acaacatacagaccacaaa 388

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pt:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: gb_vl:*
16: em_ba:*
17: em_fun:*
18: em_hum:*
19: em_in:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_ph:*
24: em_pl:*
25: em_ro:*
26: em_sts:*
27: em_sy:*
28: em_un:*
29: em_vl:*
30: em_htgo_hum:*
31: em_htgo_inv:*
32: em_htgo_rod:*
33: em_htg_hum:*
34: em_htg_inv:*
35: em_htg_rod:*
36: em_htg_other:*

pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	36.6	9.4	110000	2	LMFLCHR36_27
2	36.4	9.4	69156	2	AC090542
3	35.6	9.2	28512	2	AC013069
4	35.6	9.2	311464	3	AE003431
5	35.4	9.1	564	6	AX184941
6	35.2	9.1	153452	2	AC044895
7	35.2	9.1	202397	2	AC074394
8	35	9.0	170541	2	AC083944
9	34.2	8.8	91842	2	AC006468
10	33.8	8.7	884	2	AC074974
11	33.8	8.7	146150	2	AC068333
12	33.6	8.7	73522	2	AC006173
13	33.6	8.7	84722	9	AL133420
14	33.6	8.7	159967	2	AC048356
15	33.4	8.6	167	9	HUMPERC7
16	33.4	8.6	172	9	HUMPERCA7
17	33.4	8.6	1318	9	HUMPGCA
18	33.4	8.6	1318	11	G28579
19	33.4	8.6	1418	6	AX014878
20	33.4	8.6	74371	9	AC005369
21	33.4	8.6	81914	9	HS0J63M2
22	33.4	8.6	106711	9	AL365205
23	33.4	8.6	148042	2	AC023227
24	33.4	8.6	186092	9	AC020556
25	33.2	8.6	6741	6	E32987
26	33.2	8.6	112107	9	AL139320
27	33.2	8.6	177489	2	AC092407
28	33.2	8.5	7218	6	166494
29	32.8	8.5	153201	9	AC006454
30	32.8	8.5	202028	2	AC093086
31	32.6	8.4	42265	9	AC007136
32	32.4	8.4	108280	2	AL355862
33	32.4	8.4	148915	9	AC005030
34	32.4	8.4	151805	2	AC013741
35	32.4	8.4	158408	2	AC013336
36	32.4	8.4	176238	2	AC036122
37	32.4	8.4	182975	2	AC079085
38	32.2	8.3	168497	2	AC092410
39	32	8.2	1096	9	AF022229
40	32	8.2	1276	9	MEPEPC
41	32	8.2	1456	6	AX073600
42	32	8.2	1456	14	AF178654
43	32	8.2	32110	3	CEFA0612
44	32	8.2	50579	9	HS1086L22
45	32	8.2	99367	2	AL358940

ALIGNMENTS

RESULT 1
LMFLCHR36_27

Sequence split into 36 fragments LOCUS LMFLCHR36 Accession AL499624

Fragment Name	Begin	End
LMFLCHR36_00	1	110000
LMFLCHR36_01	100001	210000
LMFLCHR36_02	200001	310000
LMFLCHR36_03	300001	410000
LMFLCHR36_04	400001	510000
LMFLCHR36_05	500001	610000
LMFLCHR36_06	600001	710000
LMFLCHR36_07	700001	810000
LMFLCHR36_08	800001	910000
LMFLCHR36_09	900001	1010000
LMFLCHR36_10	1000001	1110000
LMFLCHR36_11	1100001	1210000
LMFLCHR36_12	1200001	1310000
LMFLCHR36_13	1300001	1410000
LMFLCHR36_14	1400001	1510000


```
KDLWEKFKLCTENVTITKSGROMFPOMKFRVSGLDAAKAYILLLDIYAADDYRYKFHN
SRMNVAKADEEMKRMV IHPDSTPTGTGOMQKVPFHKLKTNNISQKHGFEVSTTL
NSMKHYOPRHLVANDIILKLPSTFTTYVKEKEFTAVTAYONEKITQOLKTDNNPRA
KGRFDGAGKREKNCYIQAOLMSNRGSDSKINPTHVSSSRAPLHLGHAGRPHLHPHA
ALDNOQDDDDKLLDVGPPQSPPLPLSHSLQOMHAHQHSAALAWENHLGAGAGAS
EHAHAANAASAEADLRRRLQADADVERSDSDSCSVSGSTGAFRPTSTGPKRA
VGAAMAAAAGLNGGGSYSPSNISVGPRIHPSPHLLPYLYPHGLYPPPHGLLHNPA
AAAMSPAGLIPGLIFMAOLALAAOHAFGLHAYAAAGHMPVSPLOGLKSRRSPYSL
PGLSGAFDAVTPTGSNNNRSGDPPRGSGGGLGCGGVENGPRSLSSSPRRASHPPT
RPIJMSPTTPPSLHKOPRGAGAGVAAOSQHSPELSMEKMNGLVQHNGSAAAAA
AALQAEAAQHNNHTQHNNHQOQOQHQQHQQOPAPPHQHQLHSHHGATTCG
TDQYGGCGRGSGSGSG"
complement(join(47962..48065,48215..>48506))
/mrna
/gene="CG12686"
/product="CT35553"
/db_xref="FLYBASE:FBan0012686"
/db_xref="FLYBASE:FBgn0040909"
complement(<47962..>48506)
/gene="CG12686"
/map="4C2-4C2"
/db_xref="FLYBASE:FBan0012686"
/db_xref="FLYBASE:FBgn0040909"
complement(join(47962..48065,48215..48233))
/gene="CG12686" gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0012686"
/db_xref="FLYBASE:FBgn0040909"
/protein_id="AAFA5947.1"
/db_xref="GI:7290494"
/translation="MOTDSTLQEEFFGDICPLTHOPASAHCISSKTSRYPLSM"
join(<126141..126277,126550..>126709)
/gene="CG12685"
/product="CT35551"
/db_xref="FLYBASE:FBan0012685"
/db_xref="FLYBASE:FBgn0040910"
/evidence=not_experimental
<126141..>126709
/gene="CG12685"
/map="4C3-4C4"
/db_xref="FLYBASE:FBan0012685"
/db_xref="FLYBASE:FBgn0040910"
/evidence=not_experimental
join(126141..126277,126550..126709)
/gene="CG12685"
/note="CG12685 gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0012685"
/db_xref="FLYBASE:FBgn0040910"
/evidence=not_experimental
/protein_id="AAFA5948.1"
/db_xref="GI:7290495"
/translation="MDDGAASRAPETGIMDVRCRMODAGCMLAEQCMTKLLLDGL
SCAASAPLATAOLTRTROIAPAPPTAYRLRPAPPAQOQHHPOMAAVGVK"
join(140773..141143,141852..143220,143398..143553,
143862..143739,143803..143885,144067..>144128)
/gene="CG3556"
/product="CT11944"
/db_xref="FLYBASE:FBan0003556"
/db_xref="FLYBASE:FBgn0029708"
<140773..>144128
/gene="CG3556"
/map="4C3-4C4"
/db_xref="FLYBASE:FBan0003556"
/db_xref="FLYBASE:FBgn0029708"
join(141852..143220,143398..143593,143662..143739,
143803..143885,144067..144128)
/gene="CG3556" gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0003556"
/db_xref="FLYBASE:FBgn0029708"
/protein_id="AAFA5949.1"
```

```
/db_xref="GI:7290496"
/translation="MLSLSSPPMLLLVLPFANGSATVVSIAENVTLATLPSNPIS
VSEFTSSVAPISTELEENQPASSNNTLNNEADDEPTGNSNNTSSSHRNINLPSS
NSNIDNANSSSVSLSTTTSAISGVDOQDSLILAPTLANGSTPEHQITQAPP
TKGQGEAILRALDMLKERRASDYGMGDTHVYIILKELSGRDPNDSDVGHVOYIEL
EDTSLSEMEIEILAMDRLHHTLPKPDIDKLARYIALSLICDPKPHFNGHVALTL
OHHPRADIDIEALTLTISACSAAHVRROIIRLLIDISGTYDOSVDVAIAVYIALRCL
VYDHRHILHOFHVRPRGLATLIDODGSCGSLRSTALMAQLODLEVDPAGHNNRTA
ASRTILSRQRDGCWSEBPDODCEPDICVGLTIDITLALCKELCAVRLQCHVIR
ESSDPTENGEPKLVAVPGLSSSAESDSKNTSYTYTLMVGSNVESFSLVSPKNTS"
Query Match 9.2%; Score 35.6; DB 3; Length 311464;
Best Local Similarity 54.6%; Pred. No. 3.1;
Matches 71; Conservative 0; Mismatches 59; Indels 0; Gaps 0;
OY 244 tacgcaaaactatggggatcacatcattgtagcactagtcctgcgtgtagtgcacaa 303
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 99101 TACGACACGTCMAAGGCTTAATATCAGTGTGGTTAAAGTACTTGGCTATTTGTCCAA 99042
OY 304 tgggtcaatcctggttacacaattgtagcaacaatcgatcgtgcgtacacccaattt 363
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 99041 TCGTAAATACAAAGGTACACGTATTTGTTAAAGCACCTTAACATTCATTTCCACTAG 98982
OY 364 ccaaacacaa 373
|| | | | | |
Db 98981 CTATTAATAFA 98972
RESULT 5
AX184941 AX184941 564 bp DNA PAT 06-AUG-2001
LOCUS Sequence 636 from Patent WO0142467.
DEFINITION AX184941
ACCESSION AX184941
KEYWORDS AX184941.1 GI:15136315
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 564)
AUTHORS Schlegel, R., Deeds, J., Berger, A. and Zhao, X.
TITLE Genes, compositions, kits, and methods for identification,
assessment, prevention, and therapy of cervical cancer
JOURNAL Patent: WO 0142467-A 636 14-JUN-2001;
Millennium Predictive Medicine, Inc. (US)
FEATURES
source 1..564
/jorganism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 157 a 128 c 138 g 123 t 18 others
ORIGIN
Query Match 9.1%; Score 35.4; DB 6; Length 564;
Best Local Similarity 60.7%; Pred. No. 4.4;
Matches 51; Conservative 0; Mismatches 33; Indels 0; Gaps 0;
OY 259 gggatcacatcattgtagcactagtcctgcgtgtagtgcacaaatgcattcctg 318
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 435 GGGCTAATCATCTGGGNCATAGCGTTGTTCCGNGNGTGAATAATGNTTAATNCCG 494
OY 319 tacacaattgtagcaacaatcga 342
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 495 TTCACAAATTCACCAACCAACCA 518
RESULT 6
AC044895 AC044895 153452 bp DNA HTG 25-JUN-2000
LOCUS Homo sapiens chromosome 4 clone RP11-654J13 map 4, WORKING DRAFT
DEFINITION AC044895
SEQUENCE, 17 unordered pieces.
ACCESSION AC044895
VERSION AC044895.2 GI:8705180
```

KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens
ORGANISM
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS 1 (bases 1 to 153452)
TITLE Homo sapiens chrXosome 4, clone RP11-654J13
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 153452)
AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F.,
Boguslavsky, L., Boucknight, B., Brown, A., Burkett, G.,
Campbell, A., Castle, A., Choquet, Y., Colangelo, M., Collins, S.,
Collamore, A., Cooke, P., DeArrellano, K., Dewar, K., Diaz, D.,
Dodge, S., Domingo, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D.,
Galan, J., Gardina, S., Ginde, S., Goyette, M., Graham, L.,
Grand-Pierre, N., Grant, G., Hages, B., Heaford, A., Horton, L.,
Howland, J., C., Illey, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
Klein, J., Laroque, K., Lamazares, R., Landers, T., Lehocsky, J.,
Levine, R., Lieu, C., Liu, G., Locke, K., MacDonald, P., Margulis, N.,
McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R.,
Meldrum, J., Menes, L., Mihova, T., Mirand, C., Mieng, V., Morrow, J.,
Murphy, T., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,
O'Neill, D., Oliver, T. M., Oliver, J., Peterson, K., Pierre, N.,
Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Teste, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J.,
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J.,
Young, G., Zainoun, J., Zimmer, A. and Zody, M.

TITLE
JOURNAL
COMMENT
Submitted (12-APR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jun 25, 2000 this sequence version replaced gi:7543863.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www.seq.wi.mit.edu

Contact: sequence_submissions@genome.wi.mit.edu
Project Information
Center project name: L9618
Center clone name: 654_J_13

----- Summary Statistics

Sequencing vector: M13; M7815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 142959 bases at least Q40
Consensus quality: 148595 bases at least Q30
Consensus quality: 150659 bases at least Q20
Insert size: 180000; agarose-fp
Quality coverage: 3.9 in Q20 bases; agarose-fp
Quality coverage: 4.6 in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently
* consists of 17 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1
* 1177 1276: gap of 100 bp
* 1277 2611: contig of 1335 bp in length
* 2612 2711: gap of 100 bp
* 2712 3339: contig of 628 bp in length
* 3340 3439: gap of 100 bp
* 3440 5357: contig of 1918 bp in length

* 5358 5457: gap of 100 bp
* 5458 7572: contig of 2115 bp in length
* 7573 7672: gap of 100 bp
* 7673 10775: contig of 3103 bp in length
* 10776 10875: gap of 100 bp
* 10876 13352: contig of 2477 bp in length
* 13353 13452: gap of 100 bp
* 13453 17169: contig of 3717 bp in length
* 17170 17269: gap of 100 bp
* 17270 20719: contig of 3450 bp in length
* 20720 20819: gap of 100 bp
* 20820 25578: contig of 4759 bp in length
* 25579 25678: gap of 100 bp
* 25679 32478: contig of 6800 bp in length
* 32479 32578: gap of 100 bp
* 32579 38908: contig of 6330 bp in length
* 38909 39008: gap of 100 bp
* 39009 48521: contig of 9513 bp in length
* 48522 48621: gap of 100 bp
* 48622 60786: contig of 12165 bp in length
* 60787 60886: gap of 100 bp
* 60887 74000: contig of 13114 bp in length
* 74001 74100: gap of 100 bp
* 74101 95823: contig of 21723 bp in length
* 95824 95923: gap of 100 bp
* 95924 153452: contig of 57529 bp in length.

FEATURES
source
1. 153452
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="4"
/map="4"
/clone="RP11-654J13"
/clone_lib="RPCT-11 Human Male BAC"

misc_feature
1. 1176
/note="assembly-fragment"

misc_feature
1277..2611
/note="assembly-fragment"
clone_end:SP6
vector_side:right
vector_side:left
2712..3339
/note="assembly-fragment"

misc_feature
3440..5357
/note="assembly-fragment"

misc_feature
5458..7572
/note="assembly-fragment"

misc_feature
7673..10775
/note="assembly-fragment"

misc_feature
10876..13352
/note="assembly-fragment"

misc_feature
13453..17169
/note="assembly-fragment"

misc_feature
17270..20719
/note="assembly-fragment"

misc_feature
20820..25578
/note="assembly-fragment"

misc_feature
25679..32478
/note="assembly-fragment"

misc_feature
32579..38908
/note="assembly-fragment"

misc_feature
39009..48521
/note="assembly-fragment"

misc_feature
48622..60786
/note="assembly-fragment"
60887..74000
/note="assembly-fragment"
74101..95823
/note="assembly-fragment"
95924..153452
/note="assembly-fragment"
BASE COUNT 44037 a 29500 c 30194 g 48117 t 1604 others


```

misc_feature      /note="assembly_name:Contig23"
misc_feature      /note="assembly_name:Contig24"
misc_feature      /note="assembly_name:Contig25"
misc_feature      /note="assembly_name:Contig26"
clone_end:SP6
vector_side:left"
misc_feature      /note="assembly_name:Contig27"
misc_feature      /note="assembly_name:Contig28"
misc_feature      /note="assembly_name:Contig29"
BASE COUNT      62861 a 38977 c 39217 g 59228 t 2114 others
ORIGIN

```

```

Query Match
Best Local Similarity 9.1%; Score 35.2; DB 2; Length 202397;
Matches 61; Conservative 0; Mismatches 43; Indels 0; Gaps 0;
QY 250 aagcctatggggtacacactagtgatagcagctagctcgtgtagtacaatggctc 309
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 151231 AAGCTATTGTTGTCATGATATTGTTGTTACATGATGATATTCTTTAGTGGGATTTGTG 151172
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 310 aatacctgtgacacaaatgtgcgaacaacgatactgacgcat 333
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 151171 GGAACCTGTGTCACCATCATCAGAGTATACATGACACCAT 151128

```

```

RESULT 8
AC083944/c      AC083944 170541 bp DNA HTG 25-JUL-2001
LOCUS          Oryza sativa clone OSJNBa0047G15, WORKING DRAFT SEQUENCE, 4
DEFINITION    unordered pieces.
ACCESSION     AC083944
VERSION       AC083944.6 GI:15011679
KEYWORDS      HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE        Oryza sativa.
ORGANISM      Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Erihartoideae; Oryzaceae; Oryza.
REFERENCE 1 (bases 1 to 170541)
AUTHORS      McComble,W.R.
TITLE        Rice genomic sequence
JOURNAL      Unpublished
REFERENCE 2 (bases 1 to 170541)
AUTHORS      McComble,W.R.
TITLE        Direct Submission
JOURNAL      Submitted (07-OCT-2000) Lila Annenberg Hazen Genome Center, Cold
              Spring Harbor Laboratories, 1, Bungtown Road, Cold Spring Harbor,
              NY 11724, USA
COMMENT      On Jul 25, 2001 this sequence version replaced gi:14209715.
              ----- Genome Center
              Center: Lila Annenberg Hazen Genome Center, Cold Spring Harbor
              Laboratory
              Center code: CSHL
              Web site: http://www.cshl.org/genseq
              ----- Project Information
              Center project name: OSJNBa0047G15
              Center clone name: OSJNBa0047G15

```

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 4 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as

```

* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 111874: contig of 111874 bp in length
* 111875 112342: gap of unknown length
* 112343 150296: contig of 37954 bp in length
* 150297 150763: gap of unknown length
* 150764 164567: contig of 13804 bp in length
* 164568 165034: gap of unknown length
* 165035 170541: contig of 5507 bp in length.
FEATURES
      source
      1..170541
      /organism="Oryza sativa"
      /db_xref="taxon:4530"
      /clone="OSJNBa0047G15"
BASE COUNT      47036 a 36392 c 36240 g 49470 t 1403 others
ORIGIN

```

```

Query Match
Best Local Similarity 9.0%; Score 35; DB 2; Length 170541;
Matches 71; Conservative 0; Mismatches 60; Indels 0; Gaps 0;
QY 89 cggacaagaacactcgtgctaacgcaagctctgtgcaggtagaagctacgcggacctg 148
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 141330 CGGCGACACCCACACGCGCGCCGCAAGAACGCGCGGAGACAGTATCCGGCCGCGC 141271
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 149 agcagtgctgacccagcagagctctgcccgcgatgaaacattgagagcagcagc 208
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 141270 AGGCGGCGCCCGCCGCAAGCGCGCGCGCGCGCTGTAAGAAGTGTGTACGAGAGA 141211
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 209 gtaacgtccag 219
      ||| ||| |||
Db 141210 TCGCCGTGCAG 141200

```

```

RESULT 9
AC006468        AC006468 91842 bp DNA PRI 30-APR-1999
LOCUS          Homo sapiens chromosome 17, clone hRPC.159_F.22, complete sequence.
DEFINITION    AC006468
ACCESSION     AC006468
VERSION       AC006468.9 GI:4726137
KEYWORDS      HTG.
SOURCE        human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 91842)
AUTHORS      Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
              Baker,J., Baldwin,J., Barna,N., Beckely,R., Benn,J., Brown,A.,
              Castle,A., Cerny,J., Collins,M., Collins,S., Collymore,A.,
              Cooke,P., Dearellano,K., Depayre,E., Devon,K., Dewar,K.,
              Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C.,
              Funke,R., Gage,D., Galagan,J., Gattaya,S., Gilbert,D., Grant,G.,
              Hagos,B., Heathord,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,
              Karatas,A., Lehoczeky,J., Lieu,C., Locke,R., Macdonald,P.,
              Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,
              Meldrum,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
              Naylor,J., Nijhoff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,
              Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,
              Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,
              Theisley,S., Torruella-Miller,I., Vassiliou,H., Vo,A., Wagner,A.,
              Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
              ----- Direct Submission
              Submitted (29-JAN-1999) Whitehead Institute/MIT Center for Genome
              Research, 520 Charles Street, Cambridge, MA 02141, USA
              -----
              3 (bases 1 to 91842)
              Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,

```


Baker, J., Baldwin, J., Barna, N., Beckerly, R., Benn, J., Brown, A.,
 Castle, A., Cerny, J., Colangelo, M., Collins, S., Collimore, A.,
 Cooke, P., Dearellano, K., Depayre, E., Devon, K., Dewar, K.,
 Donelan, L., Doyle, M., Ferreira, P., Fitzhugh, W., Forrest, C.,
 Funke, R., Gage, D., Galagan, J., Gardyna, S., Gilbert, D., Grant, G.,
 Hagos, B., Heatford, A., Horton, L., Howland, J. C., Jones, C., Kann, L.,
 Karatas, A., Lehoczy, J., Lieu, C., Locke, K., Macdonald, P.,
 Marquis, N., McEwan, P., McGurk, A., McKernan, K., McLaughlin, J.,
 Meidrim, J., Molla, M., Morris, W., Morrow, J., Mychaleckyj, J.,
 Naylor, J., Nilloff, M., O'Connor, T., O'Donnell, P., Pavlin, B.,
 Peterson, K., Pollara, V., Riley, R., Roberts, D., Roy, A., Severy, P.,
 Stange-Thomann, N., Stojanovic, N., Stone, C., Subramanian, A.,
 Tesfaye, S., Torruella-Miller, I., Vassiliev, H., Vo, A., Wagner, A.,
 Wheeler, J., Wu, X., Wyman, D., Ye, W. J. and Zody, M.

TITLE

JOURNAL

COMMENT

Submitted (30-APR-1999) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Apr 30, 1999 this sequence version replaced gi:4713975.
 All repeats were identified using RepeatMasker: Smit, A.F.A. &
 Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Only the first 91842 base pairs of this clone are being submitted.
 The remainder overlaps accession number AC004477 (WICGR project
 L286).

FEATURES

SOURCE

location/Qualifiers
 1. 91842
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="hrpc.159_F_22"
 /clone_lib="Peter de Jong/ human PAC library"
 /map="17"
 /chromosome="17"
 992..1023
 /rpl_family="GA-rich"
 3700..3905
 /rpl_family="MIR"
 4745..4766
 /rpl_family="(CCA)n"
 7979..7999
 /rpl_family="AT-rich"
 complement(8016..8320)
 /rpl_family="AluIo"
 complement(8875..9021)
 /rpl_family="L2"
 complement(9131..9433)
 /rpl_family="AluSg"
 9451..9479
 /rpl_family="(CA)n"
 9621..9742
 /rpl_family="L2"
 9810..10123
 /rpl_family="AluSp"
 10208..10504
 /rpl_family="AluSx"
 11398..11488
 /rpl_family="CT-rich"
 12526..12534
 /rpl_family="(TTTG)n"
 complement(12555..12839)
 /rpl_family="AluSg"
 14049..14352
 /rpl_family="AluIo"
 15115..15826
 /rpl_family="AluSp"
 15971..16277
 /rpl_family="L2"
 18378..18403
 /rpl_family="AT-rich"
 18949..19231
 /rpl_family="AluSc"
 19232..19271
 /rpl_family="(TAA)n"

repeat_region 20778..21095
 /rpl_family="AluIo"
 complement(22229..22535)
 /rpl_family="AluY"
 repeat_region 22617..22949
 /rpl_family="AluYa8"
 complement(23016..23318)
 /rpl_family="AluSp"
 23386..23677
 /rpl_family="AluSx"
 23680..23975
 /rpl_family="AluSx"
 24059..24199
 /rpl_family="MIR"
 24231..24275
 /rpl_family="MIR"
 24243..24282
 /rpl_family="L2"
 complement(25339..25639)
 /rpl_family="AluSx"
 25797..26082
 /rpl_family="AluSg"
 26083..26108
 /rpl_family="(CAT)n"
 complement(26133..26437)
 /rpl_family="AluYb8"
 27293..27319
 /rpl_family="(T)n"
 28155..28189
 /rpl_family="(TG)n"
 32424..32445
 /rpl_family="AT-rich"
 34286..34579
 /rpl_family="AluSg"
 complement(34706..35003)
 /rpl_family="AluSg"
 37726..38024
 /rpl_family="AluYb"
 38112..38412
 /rpl_family="AluYb"
 38719..38746
 /rpl_family="(TTTGG)n"
 complement(38747..39033)
 /rpl_family="AluSg"
 complement(39034..39380)
 /rpl_family="L2"
 complement(39488..39537)
 /rpl_family="L2"
 39538..39839
 /rpl_family="AluSg1"
 complement(39840..40127)
 /rpl_family="L2"
 40128..40447
 /rpl_family="AluSx"
 complement(40448..41238)
 /rpl_family="L2"
 41734..42027
 /rpl_family="AluSg"
 42036..42319
 /rpl_family="AluSg"
 42403..42544
 /rpl_family="AluSx"
 42545..42843
 /rpl_family="AluSx"
 42844..43008
 /rpl_family="AluSx"
 complement(43203..43409)
 /rpl_family="MIR"
 43385..43515
 /rpl_family="L2"
 complement(44174..44245)
 /rpl_family="MLT1J"
 44246..44538

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: MIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence.submissions@genome.wi.mit.edu

----- Project Information

Center project name: L10071

Center clone name: 301_L_7

----- Summary Statistics

Sequencing vector: M13; M77815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap, version 0.960731

Consensus quality: 131662 bases at least Q40

Consensus quality: 138374 bases at least Q30

Consensus quality: 141535 bases at least Q20

Insert size: 157000; agarose-fp

Insert size: 143750; sum-of-contigs

Quality coverage: 3.8 in Q20 bases; agarose-fp

Quality coverage: 4.2 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 25 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

```

1      1408: contig of 1408 bp in length
*      1409 1508: gap of 100 bp
*      1509 3382: contig of 1874 bp in length
*      3383 3482: gap of 100 bp
*      3483 6192: contig of 2710 bp in length
*      6193 6292: gap of 100 bp
*      6293 9254: contig of 2962 bp in length
*      9255 9354: gap of 100 bp
*      9355 11843: contig of 2489 bp in length
*      11844 11943: gap of 100 bp
*      11944 13616: contig of 1673 bp in length
*      13617 13716: gap of 100 bp
*      13717 17351: contig of 3635 bp in length
*      17352 17451: gap of 100 bp
*      17452 20793: contig of 3342 bp in length
*      20794 20893: gap of 100 bp
*      20894 24809: contig of 3916 bp in length
*      24810 24909: gap of 100 bp
*      24910 28434: contig of 3525 bp in length
*      28435 28534: gap of 100 bp
*      28535 32973: contig of 4439 bp in length
*      32974 33073: gap of 100 bp
*      33074 38534: contig of 5461 bp in length
*      38535 38634: gap of 100 bp
*      38635 43314: contig of 4680 bp in length
*      43315 43414: gap of 100 bp
*      43415 49541: contig of 6127 bp in length
*      49542 49641: gap of 100 bp
*      49642 56631: contig of 6990 bp in length
*      56632 56731: gap of 100 bp
*      56732 61161: contig of 4430 bp in length
*      61162 61261: gap of 100 bp
*      61262 67595: contig of 6334 bp in length
*      67596 67695: gap of 100 bp
*      67696 73214: contig of 5519 bp in length
*      73215 73314: gap of 100 bp
*      73315 81193: contig of 7878 bp in length
*      81193 81292: gap of 100 bp
*      81293 89216: contig of 7924 bp in length
*      89217 89316: gap of 100 bp
*      89317 97756: contig of 8440 bp in length
*      97757 97856: gap of 100 bp

```

```

*      97857 109268: contig of 11412 bp in length
*      109269 109368: gap of 100 bp
*      109369 121621: contig of 12253 bp in length
*      121622 121721: gap of 100 bp
*      121722 133112: contig of 11391 bp in length
*      133113 133213: gap of 100 bp
*      133213 146150: contig of 12938 bp in length.
*      Location/Qualifiers
      source
        1..146150
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /chromosome="8"
        /map="8"
        /clone="RP11-301L7"
        /clone_lib="RPCT-11 Human Male BAC"
        1..1408
        /note="assembly_fragment"
        1509..3382
        /note="assembly_fragment"
        3483..6192
        /note="assembly_fragment"
        6293..9254
        /note="assembly_fragment"
        9355..11843
        /note="assembly_fragment"
        11944..13616
        /note="assembly_fragment"
        clone_end:5p6
        vector_side:right"
        13717..17351
        /note="assembly_fragment"
        17452..20793
        /note="assembly_fragment"
        20894..24809
        /note="assembly_fragment"
        24910..28434
        /note="assembly_fragment"
        28535..32973
        /note="assembly_fragment"
        33074..38534
        /note="assembly_fragment"
        38635..43314
        /note="assembly_fragment"
        43415..49541
        /note="assembly_fragment"
        49642..56631
        /note="assembly_fragment"
        56732..61161
        /note="assembly_fragment"
        61262..67595
        /note="assembly_fragment"
        67696..73214
        /note="assembly_fragment"
        clone_end:t7
        vector_side:right"
        73315..81192
        /note="assembly_fragment"
        81293..89216
        /note="assembly_fragment"
        89317..97756
        /note="assembly_fragment"
        97857..109268
        /note="assembly_fragment"
        109369..121621
        /note="assembly_fragment"
        121722..133112
        /note="assembly_fragment"
        133213..146150
        /note="assembly_fragment"
BASE COUNT      39106 a 32795 c 32929 g 38913 t 2407 others
ORIGIN

```



```
/note="FLAM_C repeat: matches 1. .132 of consensus"
4194. .9651
repeat_region
/note="LIP4 repeat: matches -1399. .4327 of consensus"
9667. .9724
repeat_region
/note="29 copies 2 mer tt 69% conserved"
9829. .10433
repeat_region
/note="LIMC repeat: matches 1618. .2224 of consensus"
10428. .10604
repeat_region
/note="LIM4 repeat: matches 4313. .4494 of consensus"
10695. .10976
repeat_region
/note="LIMC repeat: matches 2099. .2366 of consensus"
10977. .11280
repeat_region
/note="AlusX repeat: matches 1. .304 of consensus"
11281. .11816
repeat_region
/note="LIMC repeat: matches 1507. .2099 of consensus"
12069. .12321
repeat_region
/note="LIMC repeat: matches 923. .1179 of consensus"
12349. .13051
repeat_region
/note="LIP410 repeat: matches 5438. .6165 of consensus"
13052. .13714
repeat_region
/note="LIP repeat: matches 3989. .4657 of consensus"
13740. .14104
repeat_region
/note="TRIC repeat: matches 1. .371 of consensus"
14878. .14962
repeat_region
/note="MERSA repeat: matches 9. .109 of consensus"
15274. .15329
repeat_region
/note="14 copies 4 mer ttgt 73% conserved"
15508. .15558
repeat_region
/note="L2 repeat: matches 2648. .2694 of consensus"
15602. .15663
repeat_region
/note="31 copies 2 mer ta 69% conserved"
15990. .16079
repeat_region
/note="L2 repeat: matches 2581. .2663 of consensus"
16119. .16425
repeat_region
/note="AluI repeat: matches 1. .302 of consensus"
16559. .16704
repeat_region
/note="MERSA repeat: matches 17. .186 of consensus"
16808. .16890
repeat_region
/note="L2 repeat: matches 2666. .2749 of consensus"
16941. .17002
repeat_region
/note="MIR repeat: matches 147. .261 of consensus"
17003. .17056
repeat_region
/note="L2 repeat: matches 2685. .2738 of consensus"
17590. .17897
repeat_region
/note="AlusX repeat: matches 1. .309 of consensus"
18077. .18469
misc_feature
/note="match: GSS: Em:AQ355763"
18514. .18694
repeat_region
/note="L2 repeat: matches 2248. .2441 of consensus"
19361. .19508
repeat_region
/note="MERSB repeat: matches 10. .156 of consensus"
19570. .19874
repeat_region
/note="AlusG repeat: matches 1. .305 of consensus"
20281. .20522
repeat_region
/note="MIR repeat: matches 5. .262 of consensus"
21328. .21453
repeat_region
/note="MIR repeat: matches 49. .183 of consensus"
21677. .22165
repeat_region
/note="HAI1 repeat: matches 286. .788 of consensus"
22114. .22435
misc_feature
/note="match: STS: Em:HSPF18C2"
23043. .23290
repeat_region
/note="MIR1J repeat: matches 245. .508 of consensus"
23324. .23515
repeat_region
/note="MERSA repeat: matches 2. .189 of consensus"
23619. .23901
misc_feature
/note="match: GSS: Em:AQ072551"
24120. .24258
misc_feature
/note="match: STS: Em:G06498"
25069. .25106
repeat_region
/note="19 copies 2 mer tt 84% conserved"

repeat_region 25109. .27268
/note="LIP42 repeat: matches 3989. .6144 of consensus"
27362. .27425
repeat_region
/note="4 copies 16 mer 78% conserved"
27852. .28192
repeat_region
/note="AlusX repeat: matches 1. .302 of consensus"
28428. .28720
repeat_region
/note="AlusG repeat: matches 1. .295 of consensus"
29006. .30572
repeat_region
/note="LIP47 repeat: matches 4571. .6145 of consensus"
30930. .31226
repeat_region
/note="AlusX repeat: matches 5. .301 of consensus"
31812. .33088
repeat_region
/note="HERVL repeat: matches 4481. .5757 of consensus"
33089. .33355
repeat_region
/note="MLR2 repeat: matches 1. .267 of consensus"
33354. .33465
repeat_region
/note="7 copies 16 mer 67% conserved"
33502. .33661
repeat_region
/note="MLR2 repeat: matches 349. .510 of consensus"
33704. .33741
repeat_region
/note="19 copies 2 mer ta 78% conserved"
33798. .34064
repeat_region
/note="MLR2CA repeat: matches 1. .267 of consensus"
34138. .34553
repeat_region
/note="26 copies 16 mer 57% conserved"
34142. .34565
repeat_region
/note="212 copies 2 mer at 57% conserved"
34410. .34565
repeat_region
/note="13 copies 12 mer 67% conserved"
34614. .34666
repeat_region
/note="MLR2 repeat: matches 499. .553 of consensus"
34880. .35053
repeat_region
/note="MIR repeat: matches 63. .250 of consensus"
35226. .35525
repeat_region
/note="AlusG repeat: matches 1. .299 of consensus"
37728. .38168
misc_feature
/note="match: GSS: Em:AQ35132"
37728. .38180
misc_feature
/note="match: GSS: Em:AQ180143"
38546. .38721
repeat_region
/note="FAM repeat: matches 1. .175 of consensus"
40098. .40188
repeat_region
/note="MERSA repeat: matches 58. .155 of consensus"
40102. .40226
repeat_region
/note="MERSA repeat: matches 5. .139 of consensus"
40342. .40618
repeat_region
/note="LIR16C repeat: matches 113. .387 of consensus"
40894. .40941
repeat_region
/note="3 copies 16 mer 87% conserved"
40942. .41037
repeat_region
/note="LIM42 repeat: matches 6205. .6300 of consensus"
41111. .41216
repeat_region
/note="MIR repeat: matches 354. .463 of consensus"
41217. .41579
repeat_region
/note="MIR1A2 repeat: matches 3. .374 of consensus"
41580. .41944
repeat_region
/note="MIR1F repeat: matches 10. .354 of consensus"
42083. .42197
repeat_region
/note="MIR repeat: matches 1. .131 of consensus"
42585. .42700
misc_feature
/note="match: GSS: Em:AQ111708"

Query Match 8.7%; Score 33.6; DB 9; Length 84722;
Best Local Similarity 50.6%; Pred. No. 14;
Matches 81; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

OY 227 ggagtcgagtcgctcgagcgaagcctatggygtacatccatcagtgatcagcagc 286
|| || || || || || || || || || || || || || || || || || || || ||
DB 71607 GGTGTGCGATGCTTCCTGCTCTAGGACCTGTGATATAGTAAGAAATGTTTCTGAGC 71666
OY 287 ctgcgtctagtgacaatgggtacatctctgttacacatctgtgcacaatcgtatct 346
|| || || || || || || || || || || || || || || || || || || || ||
```

Db 71667 CTCTCCGTCTTACCTTTGCACTGACCTGACCAACATCATATTAACATACAAAC 71726

QY 347 gacgcacaccacaaatttcacaaacataacagacacaa 386
 11 11111111 11 111 11111

Db 71727 AACAAATTCACCAAGTCTTAAGGAAATTTGCTCTCAA 71766

RESULT 14
 AC048356 159967 bp DNA HTG 06-MAY-2000
 AC048356/c Homo sapiens chromosome 20 clone RP11-250P20 map 20, WORKING DRAFT
 DEFINITION
 SEQUENCE 18 unordered pieces.
 AC048356
 AC048356.2 GI:7712163
 HTG: HTGS_PHASE1; HTGS_DRAFT.
 SOURCE
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE
 1 (bases 1 to 159967)
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
 Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F.,
 Boguslavsky, L., Bouknight, B., Brown, A., Burkett, G.,
 Campiolo, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S.,
 Collymore, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S.,
 Dodge, S., Domino, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D.,
 Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,
 Grand-Pierre, N., Grant, G., Hages, B., Heaford, A., Horton, L.,
 Howard, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
 Klein, J., Larocque, K., Lamazares, R., Landers, T., Lebecky, J.,
 Levine, R., Lieu, G., Liu, G., Locke, K., MacDonald, P., Margulis, N.,
 McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheters, R.,
 Melidim, J., Menus, L., Minova, T., Miranda, C., Mienna, V., Morrow, J.,
 Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
 O'Neill, D., Oliver, T.M., Oliver, J., Peterson, K., Pierre, N.,
 Pisanic, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
 Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
 Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
 Testa, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J.,
 Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,
 Young, G., Zainoun, J., Zimmer, A. and Zody, M.
 Direct Submission
 Submitted (14-APR-2000) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On May 6, 2000 this sequence version replaced gl:7549675.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center -----
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence.submissions@genome.wi.mit.edu

----- Project Information -----
 Project name: L5951
 Center project name: 250_P-20

----- Summary Statistics -----
 Sequencing vector: M13; M7815; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 151130 bases at least Q40
 Consensus quality: 155479 bases at least Q30
 Consensus quality: 157050 bases at least Q20
 Insert size: 160000; agarose-1p
 Insert size: 158267; sum-of-contents
 Quality coverage: 4.6 in Q20 bases; sum-of-contents
 Quality coverage: 4.6 in Q20 bases; sum-of-contents

* NOTE: This is a 'working draft' sequence. It currently

* consists of 18 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 1473: contig of 1473 bp in length
 * 1474 1573: gap of 100 bp
 * 1574 3513: contig of 1940 bp in length
 * 3514 3613: gap of 100 bp
 * 3614 6910: contig of 3297 bp in length
 * 6911 7010: gap of 100 bp
 * 7011 10072: contig of 3062 bp in length
 * 10073 10172: gap of 100 bp
 * 10173 14497: contig of 4325 bp in length
 * 14498 14597: gap of 100 bp
 * 14598 19734: contig of 5137 bp in length
 * 19735 19834: gap of 100 bp
 * 19835 26423: contig of 6589 bp in length
 * 26424 26523: gap of 100 bp
 * 26524 33395: contig of 6872 bp in length
 * 33396 33495: gap of 100 bp
 * 33496 41230: contig of 7735 bp in length
 * 41231 41330: gap of 100 bp
 * 41331 48383: contig of 7053 bp in length
 * 48384 48483: gap of 100 bp
 * 48484 57555: contig of 9072 bp in length
 * 57556 57655: gap of 100 bp
 * 57656 69694: contig of 12039 bp in length
 * 69695 69794: gap of 100 bp
 * 69795 81173: contig of 11379 bp in length
 * 81174 81273: gap of 100 bp
 * 81274 93212: contig of 11939 bp in length
 * 93213 93312: gap of 100 bp
 * 93313 106969: contig of 13677 bp in length
 * 106970 107089: gap of 100 bp
 * 107090 122757: contig of 15668 bp in length
 * 122758 122857: gap of 100 bp
 * 122858 140836: contig of 17979 bp in length
 * 140837 140936: gap of 100 bp
 * 140937 159667: contig of 19031 bp in length.

FEATURES

source

1. 159967
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="20"
 /map="20"
 /clone="RP11-250P20"
 /clone_lib="RPC1-11 Human Male BAC"
 1. 1473
 /note="assembly_fragment"
 misc_feature
 1574. 3513
 /note="assembly_fragment"
 misc_feature
 3614. 6910
 /note="assembly_fragment"
 misc_feature
 7011. 10072
 /note="assembly_fragment"
 misc_feature
 10173. 14497
 /note="assembly_fragment"
 misc_feature
 14598. 19734
 /note="assembly_fragment"
 misc_feature
 19835. 26423
 /note="assembly_fragment"
 misc_feature
 26524. 33395
 /note="assembly_fragment"
 misc_feature
 33496. 41230
 /note="assembly_fragment"
 misc_feature
 41331. 48383
 /note="assembly_fragment"
 misc_feature
 48384. 57555
 /note="assembly_fragment"
 misc_feature
 57556. 69694
 /note="assembly_fragment"
 misc_feature
 69695. 81173
 /note="assembly_fragment"
 misc_feature
 81174. 93212
 /note="assembly_fragment"
 misc_feature
 93213. 106969
 /note="assembly_fragment"
 misc_feature
 106970. 122757
 /note="assembly_fragment"
 misc_feature
 122758. 140836
 /note="assembly_fragment"
 misc_feature
 140837. 140936
 /note="assembly_fragment"
 misc_feature
 140937. 159667
 /note="assembly_fragment"
 misc_feature

misc_feature vector_side:left
48484..57555
/note="assembly_fragment"
misc_feature 57656..69694
/note="assembly_fragment"
misc_feature 69795..81173
/note="assembly_fragment"
misc_feature 81274..93212
/note="assembly_fragment"
misc_feature 93313..106989
/note="assembly_fragment"
misc_feature 107090..122757
/note="assembly_fragment"
misc_feature 122858..140836
/note="assembly_fragment"
misc_feature 140937..159967
/note="assembly_fragment"
BASE COUNT 48873 a 30126 c 30354 g 48911 t 1703 others
ORIGIN

Query Match 8.7%; Score 33.6; DB 2; Length 159967;
Best Local Similarity 50.6%; Pred. No. 13;
Matches 81; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

OY 227 gagagccgagtcgtcgtaagcaagctatggtggtacatcatggtgcatagctagt 286
DB 127401 GGTGTGCGAGTGTCTCTGTAGACCTGTGAATATAGTGAATTTGTTTCTGTGAGC 127342

OY 287 ctgcgtctgtagcaatggtgcatcctgttaccaatgtcgcaacaatcgatct 346
DB 127341 CTCCTCTGTCTTACTCTGTGCACTGCACCTGCACCAATCATATAAACAATACAAAAC 127282

OY 347 gacgcataccaatttccaaaacaaataaagaacca 386
DB 127281 AACAAATTCACCAAGTTCTAAAGAAAATTGTGCTTCAA 127242

RESULT 15
HUMPEPC7/c HUMPEPC7 167 bp DNA PRI 07-JAN-1995
LOCUS Human pepsinogen C gene, exon 7.
DEFINITION M18665 J03508
ACCESSION M18665.1 GI:189817
VERSION
KEYWORDS aspartic proteinase; pepsin C; pepsinogen C; progastricsin.
SEGMENT 7 of 9
SOURCE Homo sapiens DNA.
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 167)
AUTHORS Hayano,T., Sogawa,K., Ichihara,Y., Fujii-Kuriyama,Y. and Takahashi,K.
TITLE Primary structure of human pepsinogen C gene [published erratum appears in J Biol Chem 1988 Oct 5;263(28):14592]
JOURNAL J. Biol. Chem. 263 (3), 1382-1385 (1988)
MEDLINE 88087276
COMMENT Draft entry and printed copy of sequence for [1] kindly provided by K.Takahashi, 23-NOV-1987.

FEATURES
source Location/Qualifiers
1..167
/organism="Homo sapiens"
/db_xref="taxon:9606"
/tissue_type="stomach"
/map="6pter-p21.1"
<1..9
/gene="PGC"
/note="G00-119-485"
10..157
/gene="PGC"
/note="G00-119-485"
/number=7
158..>167
Intron
exon
Intron

/gene="PGC"
/note="G00-119-485"
/number=7
BASE COUNT 29 a 50 c 54 g 34 t
ORIGIN About 0.6 kb after segment 6.

Query Match 8.6%; Score 33.4; DB 9; Length 167;
Best Local Similarity 78.3%; Pred. No. 19;
Matches 65; Conservative 0; Mismatches 16; Indels 2; Gaps 2;

OY 1 actgctggccacagtgagcagagatgtgctgtgtgcatcatgtgactgaacccttac 60
DB 104 ACTGCTGGGGGACAGTGAAGAGAGAG-GTGCTGTGTCCAGCATGTGCTTGCA-ACCCCTC 47

OY 61 agaataccagggcggggccatgc 83
DB 46 AGAACACCGACCGCGAGCGCTTGGC 24

Search completed: January 17, 2002, 08:51:03
Job time: 10351 sec

...

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OH nucleic - nucleic search, using sw model

Run on: January 17, 2002, 06:46:54 ; Search time 218.99 Seconds
(without alignments)
1518.983 Million cell updates/sec

Title: US-09-823-101-8
Perfect score: 388
Sequence: 1 actgcctggccacagctgacg.....acaacatacagacccaataa 388

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N.Geneseq_1101.*
1: /SIDS2/gcgdata/geneseq/geneseqn/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	38.8	10.0	741	21	AAZ80586
2	35.4	9.1	564	22	AAH69362
3	33.4	8.6	819	20	AAK39801
4	33.4	8.6	1357	22	AAH57422
5	33.4	8.6	1358	20	AAK39788
6	33.4	8.6	1418	20	AAZ77523
7	33.2	8.6	6741	21	AAK10595
8	33	8.5	1397	21	AAZ56732
9	32	8.2	15456	22	AAK85008
10	31.8	8.2	1366	22	AAK09486
11	31.4	8.1	1255	21	AAK00028

12	31.4	8.1	1434	21	AAA99907	CDNA encoding huma
13	31.4	8.1	1434	21	AAC58623	Human PR0205 prote
14	31.4	8.1	3192	21	AAD00023	Follistatin-relate
15	31.2	8.0	96109	22	AAF28548	Genomic fragment #
16	31	8.0	10732	21	AAK10594	Gene encoding a su
17	30.8	7.9	827	21	AAZ97295	Human prostate can
18	30.8	7.9	2949	21	AAD00024	Degenerate DNA enc
19	30.8	7.9	10930	16	AAO81226	Plasmid pM16-1. C
20	30.8	7.9	10950	16	AAO81225	Plasmid pM16. Chi
21	30.6	7.9	801	19	AAV29367	Calcium ion channe
22	30.4	7.8	977	19	AAV31204	E. coli J96 pathoge
23	29.8	7.7	3266	22	AAH72600	Human cervical can
24	29.8	7.7	3266	22	AAH72698	Human cervical can
25	29.8	7.7	3266	22	AAH72723	Human cervical can
26	29.6	7.6	2617	13	AAQ029753	DNA sequence encod
27	29.6	7.6	3384	19	AAV11346	DNA encoding C-bet
28	29.6	7.6	3384	21	AAV11347	DNA encoding C-bet
29	29.6	7.6	3387	19	AAV11347	DNA encoding C-bet
30	29.6	7.6	3387	19	AAV11345	DNA encoding C-bet
31	29.6	7.6	3405	19	AAV11345	DNA encoding C-bet
32	29.6	7.6	3405	21	AAV12453	DNA encoding a C-b
33	29.6	7.6	3492	19	AAV11348	DNA encoding C-bet
34	29.6	7.6	3492	21	AAV12456	DNA encoding a C-b
35	29.6	7.6	3730	17	AAK03195	Group B streptococ
36	29.6	7.6	4200	17	AAK03190	Group B streptococ
37	29.6	7.6	4200	19	AAV11344	Group B streptococ
38	29.6	7.6	4200	21	AAV12452	DNA encoding wild
39	29.6	7.6	21591	20	AAK13047	Enterococcus faeca
40	29.4	7.6	1122	22	AAK84392	Mouse A236 polypep
41	29.4	7.6	1866	11	AAQ04292	Encodes polypeptid
42	29.4	7.6	1949	22	AAK84391	Mouse A236 polypep
43	29.4	7.6	1949	22	AAK84416	Mouse A236 variant
44	29.4	7.6	1949	22	AAK84417	Mouse A236 variant
45	29.4	7.6	1949	22	AAK84418	Mouse A236 variant

ALIGNMENTS

RESULT 1	
AAZ80586	AAZ80586 standard; cDNA; 741 BP.
ID	AAZ80586;
AC	AAZ80586;
XX	
DT	07-APR-2000 (first entry)
XX	
DE	Human colon cancer cell line SW480 cDNA clone seq ID NO:670.
XX	
KM	Human; gene expression product; diagnosis; tumour; colon cancer;
KM	colorectal adenocarcinoma; cell line SW480; cell proliferation;
KW	cytostatic; sarcoma; breast cancer; neoplasia; dysplasia;
KW	hyperplasia; ds.
XX	
OS	Homo sapiens.
XX	
FN	W09964576-A2.
XX	
PD	16-DEC-1999.
XX	
PF	09-JUN-1999; 99WO-1B01062.
XX	
PR	10-JUN-1998; 98US-0088801.
XX	
PA	(FARB) BAYER CORP.
XX	
PI	Endege WO, Steilmann KE, Astle JH, Burgess CC, Bushnell SE;
PI	Carroll F, Catino JV, Derti A, Ford DM, Lewis ME, Monahan JE;
PI	Schlegel R;
XX	
DR	WPI: 2000-087220/07.
XX	
PT	Novel nucleic acids, used to develop products for the diagnosis and

CC	the diagnosis, monitoring, research, or treatment of conditions
CC	characterised by the expression of various cancer associated antigens.
CC	The invention provides nucleic acid sequences and encoded polypeptides
CC	which are cancer associated antigen precursors expressed in human breast
CC	cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
CC	lung cancer.
XX	
SO	Sequence 819 BP; 211 A; 213 C; 238 G; 149 T; 8 other;
Query Match	8.6%; Score 33.4; DB 20; Length 819;
Best Local Similarity	78.3%; Pred. No.0.27;
Matches 65; Conservative	0; Mismatches 16; Indels 2; Gaps 2.
Oy	1 actgctgggcccacgtgacagagagtgtgctgtgcatcatgacttgacaacctac 60
Db	446 actgctgggcccacgtgacagagagtgtgctgtgcatcatgacttgctgca-accc 503
Oy	61 agaataccagggcggggccatgac 83
Db	504 agaacaccagccggagggcctgac 526
RESULT 4	
AAH57422/C	
ID	AAH57422 standard; cDNA; 1357 BP.
XX	
AC	AAH57422;
XX	
DT	10-SEP-2001 (first entry)
XX	
DE	Human stomach cell specific cDNA sequence SEQ ID NO:262.
XX	
KW	Human; tissue specific; diagnosis; brain; heart; skeletal muscle;
KW	Lung; liver; uterus; ovary; stomach; intestine; kidney; pancreas; ss;
KW	metabolic disease; developmental disease; cytostatic; immunomodulatory;
KW	neuroprotective; gene therapy; cancer; immunopathology; neuropathology.
XX	
OS	Homo sapiens.
XX	
PN	WO200132927-A2.
XX	
PD	10-MAY-2001.
XX	
PF	02-NOV-2000; 2000WO-US30396.
XX	
PR	04-NOV-1999; 99US-0163508.
XX	
PA	(INCYTE) INCYTE GENOMICS INC.
XX	
PI	Sornasse T, Sellhammer JJ, Watson GA;
XX	
DR	WPI: 2001-291057/30.
XX	
PT	New cell and tissue specific polynucleotides useful for diagnosis,
PT	prognosis or monitoring of treatments for disorders where the gene is
PT	associated with a cancer, immunopathology or neuropathology -
XX	
PS	Claim 1: Page 191; 327pp; English.
XX	
AAH57161 to AAH57576	represent cell and tissue specific polynucleotide
CC	sequences (I). (I) can have cytosstatic, immunomodulatory and
CC	neuroprotective activities, and can be used in gene therapy. (I) and
CC	proteins (II) encoded by then are used in high throughput screening
CC	assays to select DNA molecules, RNA molecules, peptide nucleic acids,
CC	imimetics, peptides, proteins, agonists, antagonists, antibodies or
CC	their fragments, immunoglobulins, inhibitors, drug compounds and
CC	pharmaceutical agents. Expression of (I) in a sample indicates the
CC	differentiation of embryonic stem cells into a tissue selected from
CC	brain, heart, kidney, liver, lung, skeletal muscle or pancreatic
CC	tissues. (I) and (II) are used to produce an expression profile that
CC	describes a metabolic or developmental process, treatment, condition,
CC	disease or disorder. The gene profile can be used for diagnosis,

CC	prognosis or monitoring of treatments and for investigating a
CC	predisposition to a disorder where the gene is associated with a
CC	cancer, immunopathology or neuropathology.
XX	
SO	Sequence 1357 BP; 255 A; 415 C; 357 G; 330 T; 0 other;
Query Match	
Best Local Similarity 78.3%; Pred. No. 0.34;	
Matches	65; Conservative 0; Mismatches 16; Indels 2; Gaps 2
QY	1 actgctggagccacagtgaacagagatgctctgttcacatcgtgacttgaacacctac 60
DB	912 ACTGCTGGGGGACACAGTGAAGAGAG-GTCCCTGTGTCCAGATGGCTGGCA-ACCCTC 855
QY	61 agaataccagggcgggcgccatgc 83
DB	854 AGAACACGACCGCGAGGCGCTGCG 832
RESULT 5	
AAK39788/C	
ID	AAK39788 standard; DNA; 1358 BP.
XX	
AC	AAK39788;
XX	
DT	02-JUL-1999 (first entry)
XX	
DE	Gastric cancer associated gene.
XX	
KW	Cancer associated antigen; diagnosis; research; treatment; human;
KW	breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
KW	prostate cancer; ss.
XX	
OS	Homo sapiens.
XX	
PN	WO9904265-A2.
XX	
PD	28-JAN-1999.
XX	
PF	15-JUL-1998; 98WO-US14679.
XX	
PR	22-JUN-1998; 98US-0102322.
PR	17-JUL-1997; 97US-0896164.
PR	10-OCT-1997; 97US-0061599.
PR	10-OCT-1997; 97US-0061765.
PR	10-OCT-1997; 97US-0948705.
PR	11-OCT-1997; 97GB-0021697.
XX	
PA	(LUDWIG) LUDWIG INST CANCER RES.
PI	
PI	Chen Y, Gout I, Gure A, O'Hare M, Obata Y, Old LJ;
PI	Pfeundschuh M, Sahin U, Scanlan MJ, Stockert E;
PI	Tureci O;
XX	
DR	WPI: 1999-132448/11.
XX	
PT	New isolated cancer associated nucleic acids and polypeptides -
PT	isolated using sera from cancer patients, used to develop products
PT	for the diagnosis, monitoring or treatment of cancers
XX	
PS	Claim 67; Page 548; 787pp; English.
XX	
CC	The invention relates to a method for diagnosing a disorder characterised
CC	by expression of a human cancer associated antigen precursor coded for by
CC	a nucleic acid molecule (NAM). The method comprises: (a) contacting a
CC	biological sample isolated from a subject with an agent that specifically
CC	binds to the NAM, an expression product or a fragment of an expression
CC	product complexed with an HLA molecule; and (b) determining the
CC	interaction between the agent and the NAM or the expression product as a
CC	determination of the disorder. The products and methods can be used in
CC	the diagnosis, monitoring, research, or treatment of conditions
CC	characterised by the expression of various cancer associated antigens.

Db 3642 tacaccaaccaatcacacataacaa 3669
 ||||| ||| || || |||||
 RESULT 10
 AAD094486/C
 ID AAD094486 standard; cDNA: 1366 BP.
 XX
 AC AAD094486:
 XX
 DT 10-SEP-2001 (first entry)
 XX
 DE Human gastricsin cDNA from clone 1658285.
 XX
 KW Human: aspartyl protease; ASP: Alzheimer's disease; gene therapy;
 KW amnesia; anxiety; antidiabetic; neuroleptic; cancer; antifertility;
 KW metabolic; impotence; haemostatic; endometriosis; autoimmune disorder;
 KW Huntington's disease; Down syndrome; cerebral palsy; hepatitis; AIDS;
 KW seasonal affective disorder; SAD; ovulatory defect; Wilson's disease;
 KW cirrhosis; acquired immune deficiency syndrome; transgenic animal;
 KW epilepsy; emesis; gastricsin; ss.
 XX
 OS Homo sapiens.
 XX
 FH key location/Qualifiers
 FT 51..1217
 FT CDS /*tag= a
 XX /product= "Human gastricsin"
 XX
 PN MO200146398-A2.
 XX
 PD 28-JUN-2001.
 XX
 PF 12-DEC-2000; 2000MO-US33743.
 XX
 PR 22-DEC-1999; 99US-0470954.
 XX
 PA (INCY-) INCYTE GENOMICS INC.
 XX
 PI Kaser MR, Cocks BG, Loring J, Tang YT, Yue H;
 XX
 DR MPI: 2001-418060/44.
 DR P-PSDB; AAED04798.
 XX
 PT Novel mammalian aspartyl proteases useful for characterizing,
 PT diagnosing, treating, preventing Alzheimer's disease and down syndrome
 PT associated with altered expression of the aspartyl protease
 XX
 PS Disclosure: Page 97-98; 102pp; English.
 XX
 CC The present sequence is human gastricsin cDNA. Gastricsin shows
 CC structural and chemical similarity with human aspartyl protease (ASP).
 CC The invention relates to mammalian aspartyl proteases and nucleic
 CC molecules encoding them. The pharmaceutical composition comprising
 CC ASP is useful for treating diseases such as Alzheimer's disease and
 CC Down's syndrome which are associated with altered expression of ASP.
 CC ASP cDNA is useful for producing mammalian model system. ASP sequences
 CC are useful for characterisation, diagnosis, prevention and treatment
 CC of conditions such as epilepsy, Alzheimer's disease, Pick's disease,
 CC Huntington's disease, dementia, retinitis pigmentosa, multiple
 CC sclerosis, bacterial and viral meningitis, Creutzfeldt-Jakob disease,
 CC fatal familial insomnia, Down syndrome, cerebral palsy, myasthenia
 CC gravis, anxiety, seasonal affective disorder (SAD), akathisia,
 CC amnesia, diabetic neuropathy, Tourette's disorder, disorder of
 CC prolactin production, ovulatory defects, endometriosis, disruption
 CC of the oestrus cycle, disruption of the menstrual cycle, uterine
 CC fibroid, autoimmune disorders, ectopic pregnancy, cancer of the
 CC breast, galactorrhoea, disruption of spermatogenesis, abnormal sperm
 CC physiology, cancer of the testis and prostate, prostatitis, nausea,
 CC gastritis, impotence, dysphagia, indigestion, gastric carcinoma,
 CC anorexia, emesis, gastroparesis, pyrosis, cholecystitis, cirrhosis,
 CC Crohn's disease, whipple's disease, gastrointestinal haemorrhage,
 CC acquired immune deficiency syndrome (AIDS), hepatitis, jaundice,

CC Wilson's disease, intrahepatic cholestasis of pregnancy, adenomas
 CC hepatic tumours, and carcinomas. ASP cDNA sequences are used in
 CC hybridisation and amplification technologies to distinguish similar
 CC molecules in a sample. The molecules may be used to mimic human
 CC conditions, diseases, or disorders, to produce transgenic animal
 CC models for these conditions, to monitor animal toxicology studies,
 CC clinical trials, and patient treatment profiles. ASP hybridisation
 CC probes are useful in mapping naturally occurring genomic sequence.
 CC ASP cDNA is also useful in gene therapy.
 XX
 SQ Sequence 1366 BP; 267 A; 415 C; 357 G; 327 T; 0 other;
 XX
 Query Match 8.2%; Score 31.8; DB 22; Length 1366;
 Best Local Similarity 77.1%; Pred. No. 1.2;
 Matches 64; Conservative 0; Mismatches 17; Indels 2; Gaps 2;
 OY 1 actctgtggccacaagtgcagagagtgctgtgtgcatcatgagctgaacacctac 60
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 912 ACTGCTGGGCGACAGTTAGCAGAGAG-GTGCCTGTGTCCAGATGGCGCTGCA-ACCCCTC 855
 OY 61 agaataccaaggcgggcgccatgc 83
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 854 AGAACACCGACCGAGCGCTGCG 832
 XX
 RESULT 11
 AAD000028
 ID AAD000028 standard; cDNA: 1255 BP.
 XX
 AC AAD000028:
 XX
 DT 24-JUL-2000 (first entry)
 XX
 DE 3' end of human zfst2 cDNA.
 XX
 KW Follistatin-related protein; zfst2; Transforming growth factor-beta;
 KW TGF-beta family; acute inflammatory condition; bursitis; myositis;
 KW chronic inflammatory demyelinating polyneuropathy; contact dermatitis;
 KW contact vulvovaginitis; sepsis; ulcerative colitis; dermatological;
 KW antiinflammatory; antiulcer; antibacterial; immunosuppressive; ss.
 XX
 OS Homo sapiens.
 XX
 PN MO200022126-A1.
 XX
 PD 20-APR-2000.
 XX
 PF 05-OCT-1999; 99MO-US23179.
 XX
 PR 15-OCT-1998; 98US-0173044.
 XX
 PA (ZYMO) ZYMOGENETICS INC.
 XX
 PI Conklin DC, Ellsworth JL;
 XX
 DR MPI: 2000-317983/27.
 XX
 PT New polypeptide comprising follistatin homology domain is useful for
 PT treating acute inflammatory conditions, bursitis, contact dermatitis,
 PT contact vulvovaginitis, myositis, sepsis and ulcerative colitis
 XX
 PS Example 1; Page 119; 125pp; English.
 XX
 CC The present sequence is the 3' end of human follistatin-related protein
 CC zfst2 cDNA. This was derived on extension of EST (expressed sequence
 CC tag) sequence related to human zfst2. This EST lacked complete 5' and 3'
 CC regions. Extension reaction was carried using foetal brain, brain, spinal
 CC cord and retina MarathonM cDNA libraries as templates. zfst2 gene and
 CC its protein are useful for binding to members of the Transforming growth
 CC factor (TGF)-beta family and mediating central nervous system,
 CC reproductive, haematopoietic and bone-related activities. This is useful
 CC for treating acute inflammatory conditions, bursitis, chronic

1
.
.
.

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 17, 2002, 05:58:26 : Search time 93.06 Seconds

(without alignments)
944.267 Million cell updates/sec

Title: US-09-823-101-8

Perfect score: 388

Sequence: 1 actgcctggccacacgtgagc.....acaacataacgacacacaaa 388

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	38.8	10.0	741	4	US-09-328-111-670
2	33	8.5	7218	1	US-08-233-463-14
3	30.6	7.9	289	4	US-09-007-005-17
4	30.6	7.9	289	4	US-09-244-796-17
5	29.6	7.6	932	1	US-08-242-932-7
6	29.6	7.6	932	1	US-08-714-481-7
7	29.6	7.6	932	5	PCT-US95-06111-7
8	29.6	7.6	2617	1	US-08-430-024-1
9	29.6	7.6	2617	1	US-08-782-009-1
10	29.6	7.6	2617	3	US-09-017-302-1
11	29.6	7.6	3294	4	US-08-923-992A-7
12	29.6	7.6	3312	4	US-08-923-992A-3
13	29.6	7.6	3384	4	US-08-923-992A-5
14	29.6	7.6	3492	4	US-08-923-992A-9
15	29.6	7.6	3730	1	US-08-242-932-8
16	29.6	7.6	3730	1	US-08-714-481-8
17	29.6	7.6	3730	5	PCT-US95-06111-8
18	29.6	7.6	4200	1	US-08-242-932-1
19	29.6	7.6	4200	1	US-08-714-481-1
20	29.6	7.6	4200	4	US-08-923-992A-1
21	29.6	7.6	4200	5	PCT-US95-06111-1
22	28.8	7.4	7326	1	US-08-194-468-1
23	28.8	7.4	7344	3	US-08-961-739-1
24	28	7.2	431	3	US-08-476-705A-1
25	28	7.2	431	3	US-08-631-469B-1
26	28	7.2	972	1	US-07-915-934-1
27	28	7.2	972	1	US-08-325-743-1

28	27.6	7.1	540	1	US-08-258-026A-17	Sequence 17, Appl
29	27.6	7.1	540	5	PCT-US95-07541-17	Sequence 17, Appl
30	27.6	7.1	2260	2	US-08-788-750-1	Sequence 1, Appl
31	27	7.0	4695	2	US-08-231-193A-57	Sequence 1, Appl
32	27	7.0	4695	2	US-08-486-273A-57	Sequence 57, Appl
33	27	7.0	4695	3	US-08-940-086A-57	Sequence 57, Appl
34	27	7.0	6122	1	US-08-403-545-1	Sequence 1, Appl
35	27	7.0	6122	4	US-08-404-381-1	Sequence 1, Appl
36	26.8	6.9	30001	1	US-08-125-468-1	Sequence 1, Appl
37	26.8	6.9	30001	2	US-08-474-933-1	Sequence 1, Appl
38	26.6	6.9	11284	3	US-08-978-741-5	Sequence 1, Appl
39	26.4	6.8	2010	4	US-09-240-410-1	Sequence 5, Appl
40	26.4	6.8	2498	4	US-09-041-236-1	Sequence 1, Appl
41	26.2	6.8	2749	6	5240849-1	Sequence 1, Appl
42	26.2	6.8	4518	3	US-09-125-287-2	Sequence 2, Appl
43	26.2	6.8	12839	3	US-09-125-287-1	Sequence 1, Appl
44	26.2	6.8	33529	4	US-09-144-085-3	Sequence 3, Appl
45	26.2	6.8	56516	2	US-08-996-306-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-09-328-111-670
Sequence 670: Application US/09328111
Patent No. 6262333
GENERAL INFORMATION:
APPLICANT: Endege, Wilson O.
APPLICANT: Stelmann, Kathleen E.
APPLICANT: Astle, Jon H.
APPLICANT: Burgess, Christopher C.
APPLICANT: Bushnell, Steven E.
APPLICANT: Carroll III, Eddie
APPLICANT: Catino, Theodore J.
APPLICANT: Derfi, Adnan
APPLICANT: Ford, Donna M.
APPLICANT: Lewis, Marcia E.
APPLICANT: Monahan, John E.
APPLICANT: Schlegel, Robert
TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
FILE REFERENCE: CCD-257 (US)
CURRENT APPLICATION NUMBER: US/09/328,111
CURRENT FILING DATE: 1999-06-08
EARLIER APPLICATION NUMBER: US 60/088,801
EARLIER FILING DATE: 1998-06-10
NUMBER OF SEQ ID NOS: 850
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 670
LENGTH: 741
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc-feature
LOCATION: (1)...(741)
OTHER INFORMATION: n = A,T,C or G
US-09-328-111-670

Query Match 10.0%: Score 38.8; DB 4; Length 741;
Best Local Similarity 58.1%: Pred. No. 0.00046;
Matches 61; Conservative 0; Mismatches 44; Indels 0; Gaps 0;
OY 223 atcttgagtcgcagtcgtcgtacgcaagcattgggggtacacccaatagtcacacgc 282
DB 531 aagtcggaatcccaaccttcggnacccaagccttggcgttaatccatgggcataagc 590
OY 283 tagcttcgctgttagtgacaatgggtlcaatcattgtagtaacatt 327
DB 591 ttgtctccctggggggggaattgtgtaatnccggttcaacnaattc 635

```
RESULT 2
US-08-232-463-14/C
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMM
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: PTZ9PT-F15
US-08-232-463-14
```

```
Query Match      8.5%; Score 33; DB 1; Length 7218;
Best Local Similarity 1.6%; Pred. No. 0.15;
Matches 6; Conservative 208; Mismatches 163; Indels 0; Gaps 0;

QY 12 acagtgacgacgagtgctgtgcatcatgagctgaacacctacagaalaccag 71
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 1436 ACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1377

QY 72 gcgaggcgcatgcgcgacgacaaagcaactgctgtcgaacgaagctgtcgagtag 131
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 1376 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1317

QY 132 aagctacgagcgctgagacagctgctacacagcagcagctgcccgaatgaacat 191
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 1316 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1257

QY 192 tgcgagacgactgagcgctacgtccagctaatgtgagtcgagtcgctgacgcaaa 251
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 1256 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1197

QY 252 gctatggggtacacatcatggtgctacagctagcttcgtcgtcagtcagcaatggtc 311
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 1196 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1137
```

```
QY 312 tacctgtacacaattgccaacaaatcgatactgacgcatcaccacaattccaaaca 371
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 1136 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1077

QY 372 aacataacgaaccacaaa 388
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 1076 RRRRRRRRRRRATCGCAA 1060
```

```
RESULT 3
US-09-007-005-17
; Sequence 17, Application US/09007005B
; Patent No. 6258558
; GENERAL INFORMATION:
; APPLICANT: Szostak, Jack W.
; APPLICANT: Roberts, Richard W.
; APPLICANT: Liu, Rhee
; TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
; FILE REFERENCE: 00786/350003
; CURRENT APPLICATION NUMBER: US/09/007,005B
; CURRENT FILING DATE: 1998-01-14
; EARLIER APPLICATION NUMBER: 60/035,963
; EARLIER FILING DATE: 1997-01-27
; EARLIER APPLICATION NUMBER: 60/064,491
; EARLIER FILING DATE: 1997-11-06
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 289
; TYPE: RNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Translation template
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)..(289)
; OTHER INFORMATION: n = A,T,C or G
US-09-007-005-17
```

```
Query Match      7.9%; Score 30.6; DB 4; Length 289;
Best Local Similarity 5.6%; Pred. No. 0.22;
Matches 14; Conservative 112; Mismatches 124; Indels 1; Gaps 1;

QY 21 agagagtgctgctgtgcatcatgagctgaacacctacagaatcagggcgagc 80
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 12 atarurarcrcrururururarcrcrarururarcrcrarurgrnrnrnrnr 71

QY 81 tgcgagcagacaaagcaactgctgtaacgcaagctgtcgcaagtagaagctacgc 140
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 72 nrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnr 130

QY 141 gggcgtagcagtgctgctacacagcagcagctgcccgaatgaacatgcaaggag 200
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 131 nrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnr 190

QY 201 gactgagcgtagctcagcagctaatgtgagtcgagtcgctgctgacgcaagctatggg 260
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 191 nrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnrnr 250

QY 261 gtacacacata 271
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 251 rgrcgrcrua 261
```

```
RESULT 4
US-09-244-796-17
; Sequence 17, Application US/09244796
; Patent No. 6281344
; GENERAL INFORMATION:
; APPLICANT: Szostak, Jack W.
```


TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-714-481-7

Query Match 7.6%; Score 29.6; DB 1; Length 932;
Best Local Similarity 64.7%; Pred. No. 0.88;
Matches 44; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

OY 320 acacaattgtcgcaacaatcgatcgagcaccaccaattccaaacaacataac 379
||| ||| | ||||| ||||| ||| | ||| ||| |||
DB 605 ACAAATAATTGATGAACAACAAATGTCGATGCAATTATTAGAAATCAATTAAAC 664

OY 380 gaaccaaa 387
||| | ||
DB 665 GAAACTAA 672

RESULT 7

PCT-US95-06111-7
Sequence 7, Application PC/TUS9506111

GENERAL INFORMATION:

APPLICANT: Street address: 186 Grinter Hall

APPLICANT: City: Gainesville

APPLICANT: State/Province: Florida

APPLICANT: Country: US

APPLICANT: Postal code/Zip: 32611

APPLICANT: Phone number: 904-392-8929

APPLICANT: Fax number: 904-392-6600

TITLE OF INVENTION: Cloning of Non-IgA Fc Binding Forms of

TITLE OF INVENTION: the Group B Streptococcal Beta Antigens

NUMBER OF SEQUENCES: 8

CORRESPONDENCE ADDRESS:

ADDRESSEE: Saliwanichik & Saliwanichik

STREET: 2421 N.W. 41st Street, Suite A-1

CITY: Gainesville

STATE: FL

COUNTRY: USA

ZIP: 32606

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US95/06111

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/242,932

FILING DATE: 16-MAY-1994

ATTORNEY/AGENT INFORMATION:

NAME: Saliwanichik, David R.

REGISTRATION NUMBER: 31,794

REFERENCE/DOCKET NUMBER: UF142

TELECOMMUNICATION INFORMATION:

TELEPHONE: 904-375-8100

TELEFAX: 904-372-5800

INFORMATION FOR SEQ ID NO: 7:

SEQUENCE CHARACTERISTICS:

LENGTH: 932 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

PCT-US95-06111-7

Query Match 7.6%; Score 29.6; DB 5; Length 932;

Best Local Similarity 64.7%; Pred. No. 0.88;
Matches 44; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

OY 320 acacaattgtcgcaacaatcgatcgagcaccaccaattccaaacaacataac 379
||| ||| | ||||| ||||| ||| | ||| ||| |||
DB 605 ACAAATAATTGATGAACAACAAATGTCGATGCAATTATTAGAAATCAATTAAAC 664

OY 380 gaaccaaa 387
||| | ||
DB 665 GAAACTAA 672

RESULT 8

US-08-430-024-1
Sequence 1, Application US/08430024

PATENT NO. 5644030

GENERAL INFORMATION:

APPLICANT: Faulmann, Ervin L.

TITLE OF INVENTION: NOVEL GENE AND METHOD FOR PRODUCTION OF

TITLE OF INVENTION: AN IGA BINDING PROTEIN

NUMBER OF SEQUENCES: 2

CORRESPONDENCE ADDRESS:

ADDRESSEE: Saliwanichik & Saliwanichik

STREET: 2421 N.W. 41st Street

CITY: Gainesville

STATE: FL

COUNTRY: USA

ZIP: 32606

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/430,024

FILING DATE: 27-APR-1995

CLASSIFICATION: 530

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/677,209

FILING DATE: 29-MAR-1991

ATTORNEY/AGENT INFORMATION:

NAME: Saliwanichik, David R.

REGISTRATION NUMBER: 31,794

REFERENCE/DOCKET NUMBER: BL-1

TELECOMMUNICATION INFORMATION:

TELEPHONE: 904-375-8100

TELEFAX: 904-372-5800

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 2617 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

HYPOTHETICAL: NO

ANTI-SENSE: NO

ORIGINAL SOURCE:

ORGANISM: streptococcus agalactiae

STRAIN: DL471

IMMEDIATE SOURCE:

CLONE: pELF26

FEATURE:

NAME/KEY: sig_peptide

LOCATION: 320..430

FEATURE:

NAME/KEY: CDS

LOCATION: 320..1510

OTHER INFORMATION: /codon_start= 320

OTHER INFORMATION: /function= "binds to Fc region of human IgA"

OTHER INFORMATION: /product= "IgA binding protein"

OTHER INFORMATION: /number= 1

FEATURE:

NAME/KEY: RBS

```

; LOCATION: 307..311
; FEATURE:
; NAME/KEY: protein_bind
; LOCATION: 887..1507
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /bound_moiety= "Iga Fc"
; OTHER INFORMATION: /evidence= EXPERIMENTAL
; OTHER INFORMATION: /standard_name= "Human Iga-Fc binding"
; OTHER INFORMATION: /label= Iga-binding
US-08-430-024-1

Query Match
Best Local Similarity 64.7%; Pred. No. 1.5;
Matches 44; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

OY 320 acacaattgctgcaaacatgatactgacgcatcaccacaatttccaaaacacataac 379
    ||| || | ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 728 ACAAAATTTGATGAACAATGATCTGATGATATTAGAAATTAATCAATTAAAC 787
OY 380 gaaccaaa 387
    ||| ||| ||
Db 788 GAACTAA 795

RESULT 9
US-08-782-009-1
; Sequence 1, Application US/08782009
; Patent No. 5714334
; GENERAL INFORMATION:
; APPLICANT: Faulmann, Ervin L
; TITLE OF INVENTION: NOVEL GENE AND METHOD FOR PRODUCTION OF
; TITLE OF INVENTION: AN IGA BINDING PROTEIN
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Saliwanchik & Saliwanchik
; STREET: 2421 N.W. 41st Street
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/782,009
; FILING DATE: 07-JAN-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/430,024
; FILING DATE: 27-APR-1995
; APPLICATION NUMBER: US 07/677,209
; FILING DATE: 29-MAR-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Saliwanchik, David R
; REGISTRATION NUMBER: 31,794
; REFERENCE/DOCKET NUMBER: BL-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 904-375-8100
; TELEFAX: 904-372-5800
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2617 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Streptococcus agalactiae
```

```

; STRAIN: DL471
; IMMEDIATE SOURCE:
; CLONE: pELF26
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 320..430
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 320..1510
; OTHER INFORMATION: /codon_start= 320
; OTHER INFORMATION: /function= "binds to Fc region of human Iga"
; OTHER INFORMATION: /product= "Iga binding protein"
; OTHER INFORMATION: /number= 1
; FEATURE:
; NAME/KEY: RBS
; LOCATION: 307..311
; FEATURE:
; NAME/KEY: protein_bind
; LOCATION: 887..1507
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /bound_moiety= "Iga Fc"
; OTHER INFORMATION: /evidence= EXPERIMENTAL
; OTHER INFORMATION: /standard_name= "Human Iga-Fc binding"
; OTHER INFORMATION: /label= Iga-binding
US-08-782-009-1

Query Match
Best Local Similarity 64.7%; Pred. No. 1.5;
Matches 44; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

OY 320 acacaattgctgcaaacatgatactgacgcatcaccacaatttccaaaacacataac 379
    ||| || | ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 728 ACAAAATTTGATGAACAATGATCTGATGATATTAGAAATTAATCAATTAAAC 787
OY 380 gaaccaaa 387
    ||| ||| ||
Db 788 GAACTAA 795

RESULT 10
US-09-017-302-1
; Sequence 1, Application US/09017302
; Patent No. 6075128
; GENERAL INFORMATION:
; APPLICANT: Faulmann, Ervin L
; TITLE OF INVENTION: NOVEL GENE AND METHOD FOR PRODUCTION OF
; TITLE OF INVENTION: AN IGA BINDING PROTEIN
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Saliwanchik & Saliwanchik
; STREET: 2421 N.W. 41st Street
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/017,302
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/782,009
; FILING DATE: 07-JAN-1997
; APPLICATION NUMBER: US 08/430,024
; FILING DATE: 27-APR-1995
; APPLICATION NUMBER: US 07/677,209
; FILING DATE: 29-MAR-1991
; ATTORNEY/AGENT INFORMATION:
```

```

1 NAME: Saliwanchik, David R
2 REGISTRATION NUMBER: 31,794
3 REFERENCE/DOCKET NUMBER: BL-1
4 TELECOMMUNICATION INFORMATION:
5 TELEPHONE: 904-375-8100
6 TELEFAX: 904-372-5800
7 INFORMATION FOR SEQ ID NO: 1:
8 SEQUENCE CHARACTERISTICS:
9 LENGTH: 2617 base pairs
10 TYPE: nucleic acid
11 STRANDEDNESS: double
12 TOPOLOGY: linear
13 MOLECULE TYPE: DNA (genomic)
14 HYPOTHETICAL: NO
15 ANTI-SENSE: NO
16 ORIGINAL SOURCE:
17 ORGANISM: Streptococcus agalactiae
18 STRAIN: DL471
19 IMMEDIATE SOURCE:
20 CLONE: pELF26
21 FEATURE:
22 NAME/KEY: sig-peptide
23 LOCATION: 320..430
24 FEATURE:
25 NAME/KEY: CDS
26 LOCATION: 320..1510
27 OTHER INFORMATION: /codon_start=320
28 OTHER INFORMATION: /function="binds to Fc region of human IgA"
29 OTHER INFORMATION: /product="IgA binding protein"
30 OTHER INFORMATION: /number=1
31 FEATURE:
32 NAME/KEY: RBS
33 LOCATION: 307..311
34 FEATURE:
35 NAME/KEY: protein_bind
36 LOCATION: 887..1507
37 IDENTIFICATION METHOD: experimental
38 OTHER INFORMATION: /bound_moiety="IgA Fc"
39 OTHER INFORMATION: /evidence=EXPERIMENTAL
40 OTHER INFORMATION: /standard_name="human IgA-Fc binding"
41 OTHER INFORMATION: /label="IgA-binding"
42
43 OS-09-017-302-1

```

```

Query Match          7.6%; Score 29.6; DB 3; Length 2617;
Best Local Similarity 64.7%; Pred. No. 1.5;
Matches 44; Conservative 0; Mismatches 24; Indels 0; Gaps 0.

QY      320  acacaatgtcgcaacacatcgatctgcacatccaccacaattccaaacaaactaac 379
      ||| ||| | ||||| ||| ||||| ||| | ||| ||| ||| |||
Db       728  ACAAAATTTGATGAAACAAATGATTTCTGATGCATTATTGAAATTCAGAAATCAATTTAAC 787

QY      380  gaaccaaa 387
      ||| |||
Db       788  GAAACTAA 795

RESULT 11
US-08-923-992A-7
: Sequence 7, Application US/08923992A
: Patent No. 6280738
: GENERAL INFORMATION:
: APPLICANT: Tai, Joseph Y.
: APPLICANT: Blake, Milan S.
: TITLE OF INVENTION: No. 6280738-19A Fc Binding Forms of the Group B
: TITLE OF INVENTION: Streptococcal Beta Antigens
: NUMBER OF SEQUENCES: 34
: CORRESPONDENCE ADDRESSES:
: ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
: STREET: 1100 New York Avenue, N.W., Suite 600
: CITY: Washington
: STATE: D.C.
: COUNTRY: USA

```

```

? ZIP: 20005
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC Compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: Patentin Release #1.0, Version #1.30
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/923,992A
? FILING DATE: 05-SEP-1997
? CLASSIFICATION: 536
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 60/024,707
? FILING DATE: 06-SEP-1996
? ATTORNEY/AGENT INFORMATION:
? NAME: Esmond, Robert W.
? REGISTRATION NUMBER: 32,893
? REFERENCE/DOCKET NUMBER: 1438, 0140001/RWE
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (202) 371-2600
? TELEFAX: (202) 371-2540
? INFORMATION FOR SEQ ID NO: 7:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 3294 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: double
? TOPOLOGY: linear
? MOLECULE TYPE: CDNA
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 1..3294
? US-08-923-992A-7

```

```

Query Match          7.6%; Score 29.6; DB 4; Length 3294;
Best Local Similarity 64.7%; Pred. No. 1.6;
Matches 44; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

QY 320 acacaaatgltcgcaacacatcgatcagcagcatcacacaaattccaaacacataaac 379
    ||| ||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 316 ACAAAATATGATGAACAAATGATTCGATTCATATTAGAAATTAGAAAAATCAATTAAAC 375

QY 380 gaaccaaa 387
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 376 GAAACTAA 383

RESULT 12
US-08-923-992A-3
Sequence 3, Application US/08923992A
Patent No. 6280738
GENERAL INFORMATION:
APPLICANT: Tai, Joseph Y.
APPLICANT: Blake, Milan S.
TITLE OF INVENTION: No. 6280738-19A FC Binding Forms of the Group B
TITLE OF INVENTION: Streptococcal Beta Antigens
NUMBER OF SEQUENCES: 34
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
STREET: 1100 New York Avenue, N.W., Suite 600
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/923,992A
FILING DATE: 05-SEP-1997
CLASSIFICATION: 536
PRIOR APPLICATION DATA:

```



```

: APPLICATION NUMBER: US 60/024,707
: FILING DATE: 06-SEP-1996
: ATTORNEY/AGENT INFORMATION:
: NAME: Esmond, Robert W.
: REGISTRATION NUMBER: 32,893
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (202) 371-2600
: TELEFAX: (202) 371-2540
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 3312 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 1..3312
: US-08-923-992A-3

Query Match
Best Local Similarity 64.7%; Score 29.6; DB 4; Length 3312;
Matches 44; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

OY 330 acacatgtgcgaacatgcatactgcacatcccaatttccaaacacataac 379
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 316 ACAAAATTGATGAACAACAAATTCGATGCATTATTAGAAATCAATTAAAC 375
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 380 gaaccaaa 387
    ||| ||| ||
DB 376 GAACTAA 383

RESULT 13
US-08-923-992A-5
: Sequence 5, Application US/08923992A
: Patent No. 6280738
: GENERAL INFORMATION:
: APPLICANT: Tai, Joseph Y.
: APPLICANT: Blake, Milan S.
: TITLE OF INVENTION: No. 6280738-19A Fc Binding Forms of the Group B
: TITLE OF INVENTION: Streptococcal Beta Antigens
: NUMBER OF SEQUENCES: 34
: CORRESPONDENCE ADDRESSES:
: ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
: STREET: 1100 New York Avenue, N.W., Suite 600
: CITY: Washington
: STATE: D.C.
: COUNTRY: USA
: ZIP: 20005
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/923,992A
: FILING DATE: 05-SEP-1997
: CLASSIFICATION: 536
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 3312 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 1..3312
: US-08-923-992A-3
```

```

: LENGTH: 3384 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 1..3384
: US-08-923-992A-5

Query Match
Best Local Similarity 64.7%; Score 29.6; DB 4; Length 3384;
Matches 44; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

OY 320 acacatgtgcgaacatgcatactgcacatcccaatttccaaacacataac 379
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 301 ACAAAATTGATGAACAACAAATTCGATGCATTATTAGAAATCAATTAAAC 360
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 380 gaaccaaa 387
    ||| ||| ||
DB 361 GAACTAA 368

RESULT 14
US-08-923-992A-9
: Sequence 9, Application US/08923992A
: Patent No. 6280738
: GENERAL INFORMATION:
: APPLICANT: Tai, Joseph Y.
: APPLICANT: Blake, Milan S.
: TITLE OF INVENTION: No. 6280738-19A Fc Binding Forms of the Group B
: TITLE OF INVENTION: Streptococcal Beta Antigens
: NUMBER OF SEQUENCES: 34
: CORRESPONDENCE ADDRESSES:
: ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
: STREET: 1100 New York Avenue, N.W., Suite 600
: CITY: Washington
: STATE: D.C.
: COUNTRY: USA
: ZIP: 20005
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/923,992A
: FILING DATE: 05-SEP-1997
: CLASSIFICATION: 536
: INFORMATION FOR SEQ ID NO: 9:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 3492 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 1..3492
: US-08-923-992A-9
```


[illegible]

PT	Novel isolated nucleic acid molecules for diagnosing and treating melanoma, thyroid tumors, rectal, lung, breast and colon cancers
PR	
XX	
PS	
XX	Claim 1; Page 62; 94pp; English.
CC	The specification describes polynucleotides which are associated with melanoma and thyroid tumors. Specifically, these are MEL3, MEL7, Thy5, Thy6, Thy11, Thy14, and Thy15. The polynucleotides are useful for diagnosing and treating a patient with melanoma, thyroid tumour, rectal cancer, lung cancer, breast cancer or colon cancer. The present sequence represents a polynucleotide of the invention.
CC	
CC	
SO	Sequence 801 BP; 178 A; 198 C; 162 G; 165 T; 98 other:
XX	
QY	Query Match 12.7%; Score 80.2; DB 21; Length 801;
Db	Best Local Similarity 72.5%; Pred. No. 2.2e-05;
Matches	100; Conservative 0; Mismatches 36; Indels 0; Gaps 0;
QY	7 tttttttttttttttttttttttttttttttttttaaatttcgttgaggaacacctttt 66
Db	759 TTTTCTTT 700
QY	67 tttttttttttttttttttttttttttaaacaccoccttatatttttagcagcgaccoc 126
Db	699 TTCTTTTCTTAANGCINMOC 640
QY	127 acaaaaagggggaggcg 144
Db	639 CCNAAAAAGGGGAGNGG 622

RESULT	3	
AAD08501/c		
ID	AAD08501	standard; cDNA; 811 bp.
XX		
AC	AAD08501;	
XX		
DT	09-AUG-2001	(first entry)
XX		
DE	Human secreted protein-encoding gene 14	cDNA clone HNSA27, SEQ ID NO:24
KW	Human;	secreted protein; proliferative disorder; cancer; tumour;
KW	fetal abnormality;	developmental abnormality; haematopoietic disorder;
KW	immune system disorder;	AlDs; autoimmune disease; rheumatoid arthritis;
KW	inflammation;	allergy; neurological disorder; Alzheimer's disease;
KW	Parkinson's disease;	cognitive disorder; schizophrenia; asthma;
KW	skin disorder;	psoriasis; sepsis; diabetes; atherosclerosis;
KW	cardiovascular disorder;	angiogenic disorder; kidney disorder;
KW	gastrointestinal disorder;	pregnancy-related disorder; gene therapy;
KW	endocrine disorder;	infection; wound healing; vulnerability;
KW	cell culture;	chemotaxis; food additive; chromosome 2;
KW	binding partner	identification; ss.
XX		
OS	Homo sapiens.	
XX		
FH	Key	Location/Qualifiers
FT	CDS	37..591
FT		/*tag= a
FT		/product= "Human secreted protein precursor"
FT	sig-peptide	37..96
FT		/*tag= b
FT	mat-peptide	97..588
FT		/*tag= c
FT		/product= "Mature human secreted protein"
XX		
PN	WO200136432-A2.	
XX		
PD	25-MAY-2001.	
XX		
PF	15-NOV-2000;	2000MO-US31162.
XX		
RR	19-NOV-1999;	99US-0166415.
XX		

BR		30-JUN-2000; 2000US-0215136.
PA	(HUMA-) HUMAN GENOME SCI INC.	
PI	Ruben SM, Komatsoulis GA, Baker KP, Young PE,	
XX	WPI: 2001-343793/36.	
DR	P-PSDB; AAEO4211.	
XX		
PT	Isolated nucleic acid molecule encoding a human secreted protein is used in preventing, treating or ameliorating a medical condition -	
XX		
PS	Claim 1: Page 413; 509pp; English.	
CC		
CC	AAD08488-AAED08529 represent cDNAs corresponding to 18 human secreted protein genes, and AAE04199-AAE04239 represent the proteins they encode. AAED04240-AAED04297 represent human secreted protein fragments or variants.	
CC	The secreted proteins and their genes are useful for preventing, treating or ameliorating medical conditions, e.g., by protein or gene therapy. Pathological conditions can be diagnosed by determining the amount of the new protein in a sample or by determining the presence of mutations in the new genes. Specific uses are described for each of the 18 genes, based on the tissues in which they are most highly expressed, and include developing products for the diagnosis or treatment of proliferative disorders, cancer, tumours, foetal and developmental abnormalities, haematopoietic disorders, diseases of the immune system, AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation, allergies, neurological disorders (e.g., Alzheimer's disease, Parkinson's disease), cognitive disorders, schizophrenia, asthma, skin disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis, cardiovascular disorders, angiotensin disorders, kidney disorders, gastrointestinal disorders, pregnancy-related disorders, endocrine disorders, and infections. The proteins can also be used to aid wound healing and epithelial cell proliferation, to prevent skin ageing due to sunburn, to maintain organs before transplantation, for supporting cell culture of primary tissues, to regenerate tissues, to identify their cognate ligands or binding partners, and in chemotaxis, and can be used as a food additive or preservative to modify storage properties.	
CC	Antibodies specific for a protein of the invention can be used in alleviating symptoms associated with the disorders mentioned above, and in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked immunosorbent assay (ELISA). The present sequence represents a human secreted protein-encoding cDNA of the invention.	
SQ	Sequence 811 BP; 297 A; 169 C; 143 G; 202 T; 0 other;	
	Query Match 12.5%; Score 78.6; DB 22; Length 811; Best Local Similarity 64.3%; Pred.No.4.1e-05; Matches 137; Conservative 0; Mismatches 69; Indels 7; Gaps 1	
OY	7 tttcttcttcttcttcttcttcttcttcttcttcttcttcttgagggaacctttt 66 	
Db	TTTTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 747 	
OY	67 ttttttcttcttcttctttaaaccaccccccatttatcttttagagcgaccccc 126 	
Db	TTTTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 688 	
OY	127 acaaaaaggggggcggltgaaaacaactcatgtgccccttcgggggtcaacccaa 186 	
Db	-----AAAATTAAGTTGGGCATTGCGAAGAATAATTAATTTGAACCTTTTGAAGTAAACCAAG 634 	
OY	187 ggggagtcttcttctgatgataaaacaaacagg 219 	
Db	GATGTATTTCTTTTGAAAAAGATMAAACCAAGAGC 601 	
RESULT	4	
AAEC81052/c		
ID	AAEC81052 standard; cDNA: 1859 BP.	
XX		
AC	AAEC81052;	

[illegible]

[illegible]

```

PN WO200112775-A2.
XX
PD 22-FEB-2001.
XX
PF 16-AUG-2000; 2000WO-US22325.
XX
PR 17-AUG-1999; 99US-0149182.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ni J, Florence KA, Piscella M, Wei P, Baker RP;
BI Birse CE, Young PE, Komatsoulis GA, Moore PA, Soppet DR;
DR WPI; 2001-147550/15.
XX
P-PSDB; AAB74752.
XX
PT Nucleic acids encoding 25 human secreted polypeptides, useful for
PT preventing, diagnosing and/or treating e.g. cancers, Parkinson's
PT disease and diabetic retinopathy -
XX
PS Claim 1; Page 452; 485pp; English.
XX
CC AAFB1787 to AAFB1817 encode the human secreted proteins given in AAB7473
CC to AAB74772. Human secreted proteins can have activities based on the
CC tissues and cells they are expressed in. Example of activities include:
CC immunomodulatory; antisclerotic; dermatological; immunosuppressive;
CC anti-inflammatory; anti-HIV; immunostimulant; cytostatic; cardiac;
CC vascular; anti-angiogenic; ophthalmological; neuroprotectant; nootropic;
CC anticonvulsant; antialzheimers; antiparkinsonian; antimicrobial; and
CC vulnary. Human secreted proteins can be used in gene therapy and
CC vaccine. Human secreted protein nucleotide sequences (NAMI) and proteins
CC (PBP1) may be used in the prevention, diagnosis and treatment of diseases
CC associated with inappropriate polypeptide expression. For example, NAMI
CC and PBP1 may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patients genome
CC that affect the activity of proteins by expressing inactive proteins or
CC to supplement the patients own production of polypeptides. Disorders that
CC may be prevented, diagnosed and/or treated include immune disorders,
CC hyperproliferative disorders (e.g. cancers), cardiovascular diseases,
CC angiogenic disorders, neurological disorders, infectious diseases and/or
CC for promoting wound healing, regeneration and /or chemotaxis. AAFB1778 to
CC AAFB1786 and AAB74732 represent sequences used in the exemplification of
CC the present invention.
XX
SQ Sequence 712 BP; 262 A; 136 C; 149 G; 165 T; 0 other;

Query Match          12.2%; Score 76.8; DB 22; Length 712;
Best Local Similarity 59.7%; Pred. No. 8; le-05;
Matches 129; Conservative 0; Mismatches 87; Indels 0; Gaps 0.

OY 7 ttttttttttttttttttttttttttttttttttttttttttttttgggaaaaacctttt 66
DB |ttttttttttttttttttttttttttttttttttttttttttttttttttttttttt| 641
OY |tttttttttttttttttttttttaaaacaccccccttatattttaagcagaagcacccc 126
DB |ttttttttttttttttttttttttttttttttttttttttttttttttttttttttt| 581
OY |ttttttttttttttttttttttttttttttttttttttttttttttttttttttttt| 581
OY 127 acaaaaagggggggcgtggaanaaacactcatgttgcccttcgggggtcaaccgaa 186
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| 521
OY 580 GAATGGTGTCGTAATTGGCAGAAATGCATAACAGTTCCTCAATTCCTGACCACCATACCA 521
OY 187 gggagtagcttttttgatgataaaaaaaaaacaagaga 222
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| 485
OY 520 TAAACAGGATGTTCCTGTCGAACAAGCTAATGAGA 485

RESULT#      9
AAC63439/C
ID AAC63439 standard; cDNA, 876 BP.
XX AAC63439;

```



```

GENERAL INFORMATION:
APPLICANT: Nishikura, Kazuko
TITLE OF INVENTION: RNA Editing Enzyme and Methods of Use
TITLE OF INVENTION: thereof
NUMBER OF SEQUENCES: 39
CORRESPONDENCE ADDRESS:
ADDRESSEE: Howson and Howson
STREET: Spring House Corporate Cntr, P.O. Box 457
CITY: Spring House
STATE: Pennsylvania
COUNTRY: USA
ZIP: 19477
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/457,459
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/197,794
FILING DATE: 17-FEB-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/280,443
FILING DATE: 25-JUL-1994
ATTORNEY/AGENT INFORMATION:
NAME: Bak, Mary E.
REGISTRATION NUMBER: 31,215
REFERENCE/DOCKET NUMBER: WST49CUSA
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-540-9206
TELEFAX: 215-540-5818
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 6671 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 155..3832
US-08-459-1
Query Match 11.3%; Score 71.4; DB 1; Length 6671
Best Local Similarity 80.0%; Pred. No. 1; 3e-06;

```


1 APPLICANT: Anderson, Dirk M
 2 APPLICANT: Girl, Judith G
 3 TITLE OF INVENTION: Interleukin-15 Receptors
 4 NUMBER OF SEQUENCES: 15
 5 CORRESPONDENCE ADDRESS:
 6 ADDRESSEE: Immunex Corporation
 7 STREET: 51 University Street
 8 CITY: Seattle
 9 STATE: Washington
 10 COUNTRY: USA
 11 ZIP: 98101
 12
 13 COMPUTER READABLE FORM:
 14 MEDIUM TYPE: Floppy disk
 15 COMPUTER: Apple Macintosh
 16 OPERATING SYSTEM: Apple Operating System 7.1
 17 SOFTWARE: Microsoft Word for Apple, Version 5.1a
 18 CURRENT APPLICATION DATA:
 19 APPLICATION NUMBER: US/08/300, 903A
 20 FILING DATE: 06-SEPTEMBER-1994
 21 CLASSIFICATION: 435
 22 PRIOR APPLICATION DATA:
 23 APPLICATION NUMBER: USSN 08/236, 919
 24 FILING DATE: 06-MAY-1994
 25 CLASSIFICATION: 435
 26 ATTORNEY/AGENT INFORMATION:
 27 NAME: Perkins, Patricia Anne
 28 REGISTRATION NUMBER: 34,695
 29 REFERENCE/DOCKET NUMBER: 2822-A
 30 TELECOMMUNICATION INFORMATION:
 31 TELEPHONE: 206-587-0430
 32 TELEFAX: 206-233-0644
 33 INFORMATION FOR SEQ ID NO: 8:
 34 SEQUENCE CHARACTERISTICS:
 35 LENGTH: 1641 base pairs
 36 TYPE: nucleic acid
 37 STRANDEDNESS: single
 38 TOPOLOGY: linear
 39 MOLECULE TYPE: CDNA
 40 HYPOTHETICAL: NO
 41 ANTI-SENSE: NO
 42 FEATURE:
 43 NAME/KEY: CDS
 44 LOCATION: 3..839
 45 US-08-300-903A-8

RESULT 10
US-08-157-101A-4/c
; Sequence 4, Application US/08157101A
; Patent No. 5808032
; GENERAL INFORMATION:
; APPLICANT: KURIHARA, TATSUYA
; APPLICANT: MATSUKURA, SHIGEKADU
; APPLICANT: TSURUOKA, NOBUO
; APPLICANT: ARIMA, KENJI
; APPLICANT: NISHIHARA, TATSUO

TITLE OF INVENTION: ANTI-HBs ANTIBODY GENES AND EXPRESSION
 TITLE OF INVENTION: PLASMIDS THEREFOR
 NUMBER OF SEQUENCES: 9
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: PILLSBURY, MADISON & SUTRO
 STREET: 1100 NEW YORK AVENUE, N.W.
 CITY: WASHINGTON
 STATE: D.C.
 COUNTRY: USA
 ZIP: 20005
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/157,101A
 FILING DATE: 05-APR-1994
 CLASSIFICATION: 530
 ATTORNEY/AGENT INFORMATION:
 NAME: TITUS, MARLANA K
 REGISTRATION NUMBER: 35843
 REFERENCE/DOCKET NUMBER: 9437/204199
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 202-861-3711
 TELEFAX: 202-822-0944
 TELEX: 6714627 CUCH
 INFORMATION FOR SEQ ID NO: 4:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1066 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 OS-08-157-101A-4

```

Query Match      11.2%: Score 70.8; DB 1: Length 1066;
Best Local Similarity 79.2%; Pred. No. 1.2e-06;
Matches      84; Conservative    0; Mismatches   22; Indels     0; Gaps     0;

QY      7 ttttttttttttttttttttttttttttttttttttttttttttgggaaacctttt 66
          |||||
Db      1021 TTTT'TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 962
              |||||

QY      67 tttttttttttttttttttttttaaacaccccccttatctttt 112
          |||||
Db      961 TTTT'TTTTTTTTTTTTTTTTTTTTGCAAGATTCACTTATTTATT 916
              |||||

RESULT 11
US-08-628-417-6/C
; Sequence 6, Application US/08628417
; Patent No. 5627054
; GENERAL INFORMATION:
; APPLICANT: GILLESPIE, DAVID
; TITLE OF INVENTION: COMPETITOR PRIMER ASYMMETRIC
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: U.S. ARMY CHEMICAL AND BIOLOGICAL
; STREET: OFFICE OF THE CHIEF COUNSEL (AMSCB-GC)
; CITY: ABERDEEN PROVING GROUND
; STATE: MARYLAND
; COUNTRY: USA
; ZIP: 21010-5423
COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
```

```

: GENERAL INFORMATION:
: APPLICANT: Comer, Timothy W.
: APPLICANT: Wu, Kunsheng
: TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
: TITLE OF INVENTION: Plants
: FILE REFERENCE: 38-21(5195)B
: CURRENT APPLICATION NUMBER: US/09/865,419A
: CURRENT FILING DATE: 2001-05-29
: PRIOR APPLICATION NUMBER: US 60/208,063
: PRIOR FILING DATE: 2000-05-31
: NUMBER OF SEQ ID NOS: 54020
: SEQ ID NO 3373
: LENGTH: 381
: TYPE: DNA
: ORGANISM: Zea mays
: FEATURE:
: OTHER INFORMATION: Clone ID: LIB3236-040-P1-N1-D4
: US-09-865-419A-3373

```

Query Match	15.1%;	Score 95;	DB 5;	Length 381;
Best Local Similarity	-62.3%;	Pred. NO. 8.9e-06;		
Matches 149;	Conservative 0;	Mismatches 90;	Indels 0;	Gaps 0;

[illegible]

```

RESULT      3
US-09-865-439A-91961
; Sequence 91961, Application US/09865439A
; GENERAL INFORMATION:
; APPLICANT: Edgerton, Michael D
; APPLICANT: Hardeman, Kristine J.
; APPLICANT: La Rosa, Thomas J.
; TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(51936)B
; CURRENT APPLICATION NUMBER: US/09/865.439A
; CURRENT FILING DATE: 2001-05-29
; PRIOR APPLICATION NUMBER: US 60/207,458
; PRIOR FILING DATE: 2000-05-30
; NUMBER OF SEQ ID NOS: 119126
; SEQ ID NO 91961
; LENGTH: 567
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3607-057-Q6-N6-H9
; US-09-865-439A-91961

```

[illegible][illegible]

```

RESULT      4
US-09-696-664A-14530
: Sequence 14530, Application US/09696664A
: GENERAL INFORMATION:
: APPLICANT: Abad, Mark S.
: APPLICANT: Andersen, Scott E.
: APPLICANT: Dubois, Patrice
: APPLICANT: Mahadeo, Debbie A.
: APPLICANT: Masucci, James D.
: TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With
: TITLE OF INVENTION: Plants
: FILE REFERENCE: 38-21(51721)B
: CURRENT APPLICATION NUMBER: US/09/696,664A
: CURRENT FILING DATE: 2000-10-25
: PRIOR APPLICATION NUMBER: US 60/161,619
: PRIOR FILING DATE: 1999-10-26
: NUMBER OF SEQ ID NOS: 17472
: SEQ ID NO 14530
: LENGTH: 515
: TYPE: DNA
: ORGANISM: Zea mays
: FEATURE:
: NAME/KEY: unsure
: LOCATION: (1)..(515)
: OTHER INFORMATION: unsure at all n locations
: OTHER INFORMATION: Clone ID: LIB3279-217-06-N6-B10
US-09-696-664A-14530

```

	Query Match	14.5%	Score 91.6:	DB 5:	Length 515:
	Best Local Similarity	60.4%:	Pred. No. 2.8e-05:		
	Matches 151: Conservative	0:	Mismatches 99:	Indels	Gaps 0:
QY	7	ttttttttttttttttttttttttttttttttttaatttttttttggggaacotttt	66		
Dd	52	tctccttt	111		
QY	67	tttttttttttttttttttttaaacaccoccttatttttaagacgcacccc	126		
Dd	112	tt	171		
QY	127	acaaaaggggggcggtggaaaaaacaactcltgcccttcggggtcaacccaa	186		
Dd	172	aaaaaaaaagggggggaaaaaaaaaaaagggggggggcccttggaaaaaaaat	231		
QY	187	gggagttagttttttgatgaaaaaaacaggggaagagaataaccacacaaaagag	246		
Dd	232	gggaaaaaaaaaggggccctaataaaaaaaaagggggggaacccccccaanaat	291		
QY	247	gagggttggtg 256			
Dd	292	tctggggggg 301			

RESULT 5
US-09-865-419A-36424
; Sequence 36424, Application US/09865419A
; GENERAL INFORMATION:
; APPLICANT: Conner, Timothy W.
; APPLICANT: Wu, Kunsheng

```

? TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
? TITLE OF INVENTION: Plants
? FILE REFERENCE: 38-21(51935)B
? CURRENT APPLICATION NUMBER: US/09/865,419A
? CURRENT FILING DATE: 2001-05-29
? PRIOR APPLICATION NUMBER: US 60/208,063
? PRIOR FILING DATE: 2000-05-31
? NUMBER OF SEQ ID NOS: 54020
? SEQ ID NO 36424
? LENGTH: 586
? TYPE: DNA
? ORGANISM: Zea mays
? FEATURE:
? NAME/KEY: unsure
? LOCATION: (1)..(586)
? OTHER INFORMATION: unsure at all n locations
? OTHER INFORMATION: Clone ID: LIB3600-060-01-N6-B4
? US-09-865-419A-36424

```

[illegible][illegible]

[illegible]

RESULT	2
AL513643	
LOCUS	
DEFINITION	AL513643 499 bp mRNA EST 13-FEB-2001
ACCESSION	AL513643 LTI_NFL006_P12 Homo sapiens CDNA clone XCL0DBO01G12 3
VERSION	prime, mRNA sequence.
KEYWORDS	AL513643 AL513643.1 GI:12777137
SOURCE	EST.
ORGANISM	human.
REFERENCE	Homo sapiens
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
JOURNAL	Li,W.B., Gruber,C., Jesse,J. and Polayes,D.
COMMENT	Full-length cDNA libraries and normalization Unpublished (2001) Contact: Genoscope Genoscope - Centre National de Sequencage BP 191 91006 Evry cedex - France Email: seqref@genoscope.cns.fr, web : www.genoscope.cns.fr. Location/Qualifiers
FEATURES	I..499
source	

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="XCL0BB0012G12"
/clone_lib="L1_NFL006_P12"
/clone_type="Placenta"
/notes="Vector: pCMVSPORT 6; Site: 1: NotI; 1st strand cDNA
was primed with a NotI-Oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pCMVSPORT 6
vector. Library was normalized. Library was constructed by
Life Technologies. Contact : Feng Liang Life Technologies,
a division of Invitrogen 9800 Medical Center Drive
Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com."

```

Query Match	15.2%	Score 95.8;	DB 10;	Length 499;
Best Local Similarity	54.9%	Pred. No. 0.02;		
Matches 157;	Conservative 12;	Mismatches 117;	Indels 0;	Gaps 0;

[illegible]

RESULT	3			
LOCUS	CNS01523			
DEFINITION	CNS01523 1201 bp DNA	GSS	26-JUL-1999	
ACCESSION	Drosophila melanogaster genome survey sequence SP6 end of BAC			
VERSION	BA0N15106 of DrosBAC library from Drosophila melanogaster (fruit			
KEYWORDS	fly), genomic survey sequence.			
ORGANISM	ALI06089			
SOURCE	ALI06089.1 GI:5620019			
REFERENCE	GSS.			
AUTHORS	fruit fly.			
TITLE	Plasmid Drosophila melanogaster			
JOURNAL	Eukaryota: Metazoa: Arthropoda: Tracheata; Hexapoda; Insecta;			
COMMENT	Pterygota: Neoptera: Endopterygota; Diptera: Brachycera;			
	Muscomorpha: Ephydroidea; Drosophilidae; Drosophila.			
	1 (bases 1 to 1201)			
	Genoscope.			
	Direct Submission			
	Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :			
	Bp 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr			
	- Web : www.genoscope.cns.fr)			
	Determination of this BAC-end sequence was carried out as part of a			
	collaboration with the European Drosophila Genome Project (EDGP) -			
	http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC			
	library (Dros BAC) was made by Alain Billaud at CEPH (Centre			
	d'Etude du Polymorphisme Humain) with funding provided by a MRC			
	project grant. The DNA was prepared from embryos by Alain Bucheton			
	and Genevieve Payan. It has been constructed in the vector			
	pBeloSAC11.			

```

FEATURES
source      location/Qualifiers
            1. .1201
              /organism="Drosophila melanogaster"
              /db_xref="pbeIOBAC11"
              /db_xref="taxon:7227"
              /clone_1_id="DrosBAC"
              /clone="BACN15106"
              /note="end : Sp6"
BASE COUNT  177 a      221 c      322 g      279 t      202 others
ORIGIN

```

[illegible]

OY	127	acaaaagggggggggcggtgtaaaaaaacatctatgtggccttcggggttcaaccca	186
Db	559	AGGGAGGRCACAGAGGAGRAGRGCCGAGGGGGGGCGCCGACANAAGGGCCGAKCCCKNKBK	628
OY	187	ggagtgatgtttttttgtatgtaaaaaaacagagagtgagtgatataaccacccaataaag	246
Db	629	SKNKAANBCCCCCACKACACAMACCACCCKCCCGBMCKACBACCAACACCAARCAAC	688
OY	247	gaggggttggcagaagagggaaatccccaagaatgtccagagaagaaccccatltgtga	306
Db	689	AAGCMGGGCGGAAAAAAACGGCGAMSRGCCGCCGCCRCGRGGGGGGGGGGSCCG	748
OY	307	tgacgttccccacacacccaagcatatgtgtgtgctcacaacacatccccat	366
Db	749	GGSGGGGACMAAAMAMMGACACAAGRMGCCACAGGMAAACMAAAGCMAGRCAMG	808
OY	367	aaaaagaggaatcagtcgaagaanaaagtcacacagcgtgtgtaaccaagaaggacaac	426
Db	809	GGGAAGGGGGCGGAGRGRRRAAAMMRGGRMCMYGGAGGGGAGGGMGMAMGAMGGG	868
OY	427	accgcgtgttttgaagaaaaaacacagaggggccacacagcgcggggggggccaaag	486
Db	869	GGGGGGGGGCCSCSRMCACGAAAAAAMMVGGGGMSMCRMKMGGGGGGASGCCCCV	928
OY	487	accgaatcaacacacaggggcgcgcacaaatgtgtcccgactcacaacccaagtgggaa	546
Db	929	CASMGKRAGVVSVGRGSGSGSGSCSCCGCCCGCCSCSCSCSCSCSCSCSGCGCCRGSGV	988
OY	547	aatcgtgagcaaaaggtctcgagagagaagaatgtccccgcgcgc	588
Db	989	AMCRGCGCGSCGCGGGGGGGGGCGGGGGGGGGGGGGGGGGGGC 1030	

RESULT	4
LOCUS	A1636456
DEFINITION	A1636456 293 bp mRNA EST 26-APR-1999
ACCESSION	U61002.x1 NCI-CGAP_kid8 Homo sapiens CDNA clone IMAGE:223059 3'
VERSION	similar to contains TAK1.b2 MERR2 repetitive element ; , mRNA
KEYWORDS	sequence.
SOURCE	A1636456
ORGANISM	A1636456.1 GI:4687786
REFERENCE	EST.
AUTHORS	human.
TITLE	Homo sapiens
JOURNAL	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
COMMENT	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
	1 (bases 1 to 293)
	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap .
	National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
	Tumor Gene Index
	Unpublished (1997)
	Contact: Robert Strausberg, Ph.D.
	Email: cgapbs-remail.nih.gov
	Tissue Procurement: Christopher Moskalko, M.D., Ph.D., Michael R.
	Emmett-Buck, M.D., Ph.D.
	CDNA Library Preparation: Life Technologies, Inc.
	DNA Library Arrayed by: Greg Lennon, Ph.D.
	DNA Sequencing by: Washington University Genome Sequencing Center
	clone distribution: NCI-CGAP clone distribution information can be
	found through the I.M.A.G.E. Consortium/LNL at:
	www.bio.lnl.gov/dbfp/image/image.html
	Seq primer: -400P from Glbco
	High quality sequence step: 283.
	Location/Qualifiers
FEATURES	1..293
SOURCE	

```

FEATURES
source
location/Qualifiers
1. .293
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2233059"
/clone_1b="NCI_CGAP_K1d8"
/tissue_type="renal cell tumor"
/lab_host="DH10B"
/note=Organ: kidney; Vector: pCMV-SPORT6; Site:1; Salin

```

BASE COUNT	ORIGIN
66 a	33 c 55 g 139 t

Site 2: Notif. Cloned unidirectionally. Primer: 019g dT
Average insert size 1.2 Kb. Life Technologies catalog #: 11524-014"

[illegible]

```

RESULT      5
AI493248
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 735)
NCI-CCGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CCGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bmrp/image/image.html
Insert length: 1468 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 304.
Location/Qualifiers
1..735

```

```

FEATURES
SOURCE
Location/Qualifiers
1. .735
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2131991"
/clone_lib="NCI CGAP Gas4"
/tissue_type="poorly differentiated adenocarcinoma with
signed ring cell features"
/lab_host="DH10B"
/note="Organ: stomach; Vector: pCMV-SPORT6; Site:1; Salt;

```


[illegible][illegible]

[illegible]

SOURCE		human.
ORGANISM		Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 308)
REFERENCE		NCI-CGAP http://www.ncbi.nlm.nih.gov/hicgap . National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index Unpublished (1997) Contact: Robert Strausberg, Ph.D. Email: cgaaps-remail.nhi.gov Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: Greg Lennon, Ph.D. DNA Sequencing by: Washington University Genome Sequencing Center Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E Consortium/ LMNL.ac : www.bio.liml.gov/dbirp/image/image.html Seq primer: -400P from Gibco High quality sequence stop: 306. Location/Qualifiers 1..308
FEATURES		/organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:2287226" /clone_lib="NCI CGAP_Utl1" /tissue_type="well-differentiated endometrial adenocarcinoma, 7 pooled tumors" /lab_host="DH10B" /note="Organ: uterus; Vector: pCMV-Sport6; Site_1: SalI; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.75 kb. Life Technologies catalog #: 11538-014"
BASE COUNT	93 a	50 c 31 g 134 t
ORIGIN		
Query Match	14.2%; Score 89.6;	DB 10; Length 308;
Best Local Similarity	81.2%; Pred. No. 0.13:	
Matches :104; Conservative	0; Mismatches	24; Indels 0; Gaps 0.
Oy	7	ttgggaaaacctttt 66
Db	30	tty 89
Oy	67	ttttttttttttttttttttaaacacccccccttatcttttagagcgcaacccc 126
Db	90	ttytytytytytytytytytytaaaaaaaaaaaCCCCCTTTTTTTCCGCCCCCCC 149
Oy	127	acaacaaaa 134
Db	150	AAAAAAAAA 157
RESULT 12		
AL513871		
LOCUS	AL513871	mRNA EST 13-FEB-2001
DEFINITION	AL513871 LTL.NFL006.PL2 Homo sapiens cdna clone CU0BA002ZA07 3	
ACCESSION	AL513871	
VERSION	AL513871.1 GI:12777365	
KEYWORDS	EST.	
SOURCE	human.	
ORGANISM	Homo sapiens	
Eukarya; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
1 (bases 1 to 1101)		
Ll.W.B., Gruber,C.; Jesse,J. and Polayes,D.		
Full-length cDNA libraries and normalization		
Unpublished (2001)		
Contact: Genoscope		
Genoscope - Centre National de Sequencage		

Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center

CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/BLN at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 1901 Std Error: 0.00
Seq primer: -400P from Gibco

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 17, 2002, 06:46:48 ; Search time 218.99 Seconds
(without alignments)
422.810 Million cell updates/sec

Title: US-09-823-101-6

Perfect score: 108
Sequence: 1 accatttctatatgtgaga.....aaaaaaaaaaagcgcg 108

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: N_Geneseq_1101.*
2: /SIDS2/gcgdata/geneseq/geneseqn/NA1980.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseqn/NA1981.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseqn/NA1982.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseqn/NA1983.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseqn/NA1984.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseqn/NA1985.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseqn/NA1986.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseqn/NA1987.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseqn/NA1988.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseqn/NA1989.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseqn/NA1990.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseqn/NA1991.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseqn/NA1992.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseqn/NA1993.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseqn/NA1994.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseqn/NA1995.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseqn/NA1996.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseqn/NA1997.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseqn/NA1998.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseqn/NA1999.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	76.6	70.9	247	21	AAC04513
2	75	69.4	1133	20	AAZ24435
3	75	69.4	5145	21	AAC98196
4	61.8	57.2	982	22	AAZ21878
5	61.8	57.2	38682	22	AAZ21770
6	59.4	55.0	555	22	AAH33403
7	54.2	50.2	3265	21	AAZ65095
8	54.2	50.2	3265	22	AAAF2092
9	54.2	50.2	3265	22	AAAF4241
10	52	48.1	359	22	AAH83313
11	51.4	47.6	476	22	AAH83221

12	51.2	47.4	2802	20	AAZ56243	Human vitamin D re
13	51.2	47.4	2910	20	AAZ56242	Human vitamin D re
14	51	47.2	1052	21	AAZ63358	Human secreted pro
15	50.8	47.0	1064	21	AAZ36918	DNA encoding Xenop
16	50.2	46.5	1580	22	AAI63862	Human polynucleoti
17	50.2	46.5	1931	20	AAZ00477	Human secreted pro
18	50.2	46.5	1932	20	AAZ00411	Human secreted pro
19	50.2	46.5	1951	22	AAI63833	Human polynucleoti
20	49.8	46.1	251	22	AAI57596	Human colorectal c
21	49.2	45.6	665	22	AAH34083	Human colon cancer
22	49.2	45.6	1153	20	AAV55748	Human secreted pro
23	49	45.4	1354	21	AAZ77921	Human cancer assoc
24	49	45.4	3527	21	AAC99092	Human pancreatic c
25	49	45.4	3648	21	AAC75860	Human ORFX ORF1415
26	48.8	45.2	887	21	AAC78089	Human cancer assoc
27	48.8	45.2	2226	21	AAC77975	Human cancer assoc
28	48.8	45.2	4189	21	AAC77054	Human ORFX ORF2609
29	48.6	45.0	1355	22	AAH34544	Human colon cancer
30	48.6	45.0	1640	20	AAFI5794	Human prostate can
31	48.6	45.0	2630	22	AAH33733	Human colon cancer
32	48.4	44.8	797	21	AAC79717	Human secreted pro
33	48.4	44.8	1080	21	AAC65612	Human secreted pro
34	48.4	44.8	3138	12	AAQ11712	Shuttle vector PMU
35	48.2	44.6	772	22	AAZ32766	Human secreted pro
36	48.2	44.6	785	21	AAC77913	Human cancer assoc
37	48.2	44.6	1181	19	AAV59803	Human secreted pro
38	48.2	44.6	1212	19	AAV59686	Human secreted pro
39	48.2	44.6	3060	22	AAAD02924	Human PRO5723 CDNA
40	48.2	44.6	3060	22	AAAF4263	Human PRO5723 nucl
41	48.2	44.6	3060	22	AAC91489	Human PRO5723 CDNA
42	48.2	44.6	4257	20	AAZ98030	Nucleic acid encod
43	48	44.4	404	22	AAAF6487	Novel human polyu
44	48	44.4	894	20	AAV99723	Human adult retina
45	48	44.4	997	20	AAZ52274	Protein PRO244 CDN

ALIGNMENTS

RESULT 1
AAC04513
ID AAC04513 standard; cDNA: 247 BP.
XX
AC AAC04513;
XX
DT 06-OCT-2000 (first entry)
XX
DE Human secreted protein 5' EST, SEQ ID NO: 8588.
XX
KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping; ss.
XX
OS Homo sapiens.
XX
PN EPI033401-A2.
XX
PD 06-SEP-2000.
XX
PF 21-FEB-2000; 2000EP-0200610.
XX
PR 26-FEB-1999; 99US-0122487.
XX
XX (GEST) GENSET.
XX
PA Dumas Milne Edwards J, Duclert A, Giordano J;
XX
PI Dumas Milne Edwards J, Duclert A, Giordano J;
XX
DR WPI: 2000-500381/45.
XX
PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS Claim 1; SEQ ID 8588; 71pp + CD-ROM; English.

50 Sequence 247 BP; 88 A; 37 C; 33 G; 88 T; 1 other;

Matches 87; Conservative 1; Mismatches 0; Indels 1; Gaps 1;

Db 159 ttgtatatgtgagatgtttaataatgtgaaaaaratgaaataaagcatgtttgtt 218

Db 219 ttccaaagaaaaaaaaaaaaaa 247

ID AA224435 standard; cDNA; 1133 BP.

DT 14-FEB-2000 (first entry)

DE Human bladder tumour cDNA library derived EST 47.

KW treatment; gene therapy; EST; ss.

OS Homo sapiens.

PN DE19818619-A1.

PD 28-OCT-1999.

PF 21-APR-1998;

PR 21-APR-1998; 98DE-1018619.

PA (META-) METAGEN GES GENOMFORSCHUNG MBH.

PI Rosenthal A, Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E;

DR WPT; 1999-612028/53.

PT of therapeutic agents -

PS Claim 3; Page 96; 132pp; German.

CC directly treat this form of cancer (including expression from gene

Sequence 1133 BP; 383 A; 198 C; 179 G; 373 T; 0 other;

Matches	86;	Conservative	0;	Mismatches	0;	Indels	1;	Gaps	1;
---------	-----	--------------	----	------------	----	--------	----	------	----

Db 1040 ttgtatatgtgagatgtttaataatgtgaaaaaaaaatgaaataagcatgtttgtt 1099

Db 1100 ttccaaagaaaaaaaaaaaaaa 1126

ID AAC98196 standard; cDNA; 5145 BP.

AC AAC98196;

DT 09-MAR-2001 (first entry)

DE Human colon cancer antigen nucleotide sequence SEQ ID NO:206.

KW Human; colon cancer; colon cancer antigen; diagnosis; detection;

KW immunomodulatory; muscular; gynaecological; gastrointestinal;

KW neural disorder; immune system disorder; muscular disorder;

KW infectious disease; cardiovascular disorder; ss.

05 Homo sapiens.

PN W0200055351-A1.

PD 21-SEP-2000.

PF 08-MAR-2000; 2000WO-US05883.

PR 12-MAR-1999; 99US-0124270.

PA (HUMA-) HUMAN GENOME SCI INC.

PI Rosen CA, Ruben SM;

DR WPT; 2000-587534/55.

XX

PT antigens, useful for the treatment, prevention, and diagnosis of colon

XX

CC neuroprotective, immunomodulatory, gynaecological, gastrointestinal,

CC vulnerable, nephrotropic, anti-infective and antibacterial activities, and
 CC can be used in gene therapy. The colon cancer antigen polynucleotides,
 CC proteins and antibodies to the proteins are useful for the prevention,
 CC treatment and diagnosis of colon disorders, such as colon cancer. The
 CC polynucleotides may be used in diagnostics and research, such as for
 CC chromosome identification, and as hybridisation probes. The proteins
 CC may also be used to prevent diseases such as neural disorders, immune
 CC system disorders, muscular disorders, reproductive disorders,
 CC gastrointestinal disorders, wounds, renal disorders, infectious
 CC diseases, and cardiovascular disorders. AAC98764 to AAC98772 and
 CC AAC94007 represent sequences used in the exemplification of the present
 CC invention.

SO Sequence 5145 BP; 1099 A; 1302 C; 1499 G; 1236 T; 9 other;

Query Match 69.4%; Score 75; DB 21; Length 5145;
 Best Local Similarity 98.9%; Pred. No. 0.00031;
 Matches 86; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

OY 6 ttgtatattgagatgtttaataaa-tgtgaataaataagaatgattgtt 64
 |||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 5006 ttgtatattgagatgtttaataaaatgtgaataaataagaatgattgtt 5065

OY 65 ttccaaagaagaaaaaa 91
 |||||||||||||||||||||||||||

DB 5066 ttccaaagaagaaaaaa 5092

RESULT 4
 AAS21878
 ID AAS21878 standard; DNA; 982 BP.
 AC AAS21878:
 XX 24-OCY-2001 (first entry)
 DT Human collagen gene COL1A2 3' UTR.
 DE
 XX Human: collagen; COL1A1; COL1A2; COL9A1; COL9A2; COL9A3; ds;
 KW osteoporosis; multiple epiphyseal dysplasia; osteogenesis imperfecta;
 KM shortness of stature; low bone density; gene therapy; 3' UTR;
 KM untranslated region.
 XX Homo sapiens.
 OS
 XX US6265157-B1.
 PN 24-JUL-2001.
 PD
 XX 03-OCY-1997; 97US-0943731.
 PF
 XX 03-DEC-1991; 91US-0803628.
 PR 13-MAR-1994; 94US-0212322.
 PR
 XX (UYAL-) UNIV ALLEGHENY HEALTH SCI.
 PA (UYJE-) UNIV JEFFERSON THOMAS.
 PA (UYOU-) UNIV ODU.
 PA
 XX Prockop DJ, Spotila LD, Deltas CD, Sereda L, Westerhausen Larson A;
 PI Pack M, Collage A, Early J, Koerkhoe J, Ala-Kokko L, Annunen S;
 PI Pihlajamaa T, Vuoristo M, Paasilta P;
 XX WPI: 2001-432201/46.
 DR
 XX Detecting collagen gene alteration, useful for diagnosing osteoporosis,
 PT multiple epiphyseal dysplasia, osteogenesis imperfecta, shortness of
 PT stature and low bone density in humans -
 XX Example 4; Fig 5B; 617pp; English.
 PS The invention relates to detecting a collagen gene alteration associated
 CC with a pathological condition in a human subject by obtaining from the

CC subject a sample nucleic acid containing a portion of at least 15
 CC consecutive nucleotides from the segment of the COL1A1 gene extending in
 CC the 5' to 3' direction from 78 nucleotides of intron 27 located adjacent
 CC exon 28 through the 3' end of intron 51, where the portion contains an
 CC intronic nucleotide and a first and second site, determining the sequence
 CC of the portion and comparing the sequence of the COL1A1 gene with the
 CC corresponding consensus sequence of the COL1A1 gene where a difference
 CC between the sequence of the portion and the consensus sequence indicates
 CC the presence of the collagen alteration in the subject. The method is
 CC used for detecting abnormalities in a COL1 or COL2 gene is useful for
 CC determining whether a subject is afflicted with pathological conditions
 CC associated with an altered collagen gene such as osteoporosis, multiple
 CC epiphyseal dysplasia, osteogenesis imperfecta, shortness of stature and
 CC low bone density. Identification of an abnormality in a collagen gene is
 CC also useful for designing a therapeutic nucleotide or gene therapy agent
 CC which can be administered to the subject to correct or alleviate the
 CC abnormality. The method is useful for detecting mutations in both the
 CC coding and non-coding sequences of any of the COL1 or COL2 genes.
 CC Therefore the method can be used to detect collagen gene alterations
 CC which affect either the primary sequence of a collagen protein chain,
 CC splicing of the mRNA encoding such chains or regulation of expression of
 CC the genes encoding such chains. The present sequence is the 3' UTR
 CC (untranslated region) of a collagen gene of the invention.

SO Sequence 982 BP; 323 A; 169 C; 159 G; 331 T; 0 other;

Query Match 57.2%; Score 61.8; DB 22; Length 982;
 Best Local Similarity 90.6%; Pred. No. 0.041;
 Matches 77; Conservative 0; Mismatches 7; Indels 1; Gaps 1;

OY 6 ttgtatattgagatgtttaataaa-tgtgaataaataagaatgattgtt 64
 |||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 771 ttgtatattgagatgtttaataaaatgtgaataaataagaatgattgtt 830

OY 65 ttccaaagaagaaaaaa 89
 |||||||||||||||

DB 831 ttccaaagaacatattgagtaaa 855

RESULT 5
 AAS21770
 ID AAS21770 standard; DNA; 38682 BP.
 AC AAS21770:
 XX 24-OCY-2001 (first entry)
 DT Human gene for collagen COL1A2.
 DE
 XX Human: collagen; COL1A1; COL1A2; COL9A1; COL9A2; COL9A3; ds;
 KW osteoporosis; multiple epiphyseal dysplasia; osteogenesis imperfecta;
 KM shortness of stature; low bone density; gene therapy.
 KM
 XX Homo sapiens.
 OS
 XX US6265157-B1.
 PN 24-JUL-2001.
 PD
 XX 03-OCY-1997; 97US-0943731.
 PF
 XX 03-DEC-1991; 91US-0803628.
 PR 13-MAR-1994; 94US-0212322.
 PR
 XX (UYAL-) UNIV ALLEGHENY HEALTH SCI.
 PA (UYJE-) UNIV JEFFERSON THOMAS.
 PA (UYOU-) UNIV ODU.
 PA
 XX Prockop DJ, Spotila LD, Deltas CD, Sereda L, Westerhausen Larson A;
 PI Pack M, Collage A, Early J, Koerkhoe J, Ala-Kokko L, Annunen S;
 PI Pihlajamaa T, Vuoristo M, Paasilta P;

PR 03-JUN-1998; 98US-0087827.
PR 04-JUN-1998; 98US-0088021.
PR 04-JUN-1998; 98US-0088025.
PR 04-JUN-1998; 98US-0088028.
PR 04-JUN-1998; 98US-0088029.
PR 04-JUN-1998; 98US-0088030.
PR 04-JUN-1998; 98US-0088033.
PR 04-JUN-1998; 98US-0088326.
PR 05-JUN-1998; 98US-0088167.
PR 05-JUN-1998; 98US-0088202.
PR 05-JUN-1998; 98US-0088212.
PR 05-JUN-1998; 98US-0088217.
PR 09-JUN-1998; 98US-0088655.
PR 10-JUN-1998; 98US-0088722.
PR 10-JUN-1998; 98US-0088730.
PR 10-JUN-1998; 98US-0088734.
PR 10-JUN-1998; 98US-0088738.
PR 10-JUN-1998; 98US-0088740.
PR 10-JUN-1998; 98US-0088741.
PR 10-JUN-1998; 98US-0088742.
PR 10-JUN-1998; 98US-0088810.
PR 10-JUN-1998; 98US-0088811.
PR 10-JUN-1998; 98US-0088824.
PR 10-JUN-1998; 98US-0088825.
PR 10-JUN-1998; 98US-0088826.
PR 11-JUN-1998; 98US-0088858.
PR 11-JUN-1998; 98US-0088861.
PR 11-JUN-1998; 98US-0088863.
PR 11-JUN-1998; 98US-0088876.
PR 12-JUN-1998; 98US-0089090.
PR 12-JUN-1998; 98US-0089105.
PR 16-JUN-1998; 98US-0089440.
PR 16-JUN-1998; 98US-0089512.
PR 16-JUN-1998; 98US-0089514.
PR 17-JUN-1998; 98US-0089532.
PR 17-JUN-1998; 98US-0089538.
PR 17-JUN-1998; 98US-0089598.
PR 17-JUN-1998; 98US-0089599.
PR 17-JUN-1998; 98US-0089600.
PR 17-JUN-1998; 98US-0089653.
PR 18-JUN-1998; 98US-0089801.
PR 18-JUN-1998; 98US-0089907.
PR 18-JUN-1998; 98US-0089908.
PR 19-JUN-1998; 98US-0089947.
PR 19-JUN-1998; 98US-0089948.
PR 19-JUN-1998; 98US-0089952.
PR 22-JUN-1998; 98US-0090246.
PR 22-JUN-1998; 98US-0090252.
PR 22-JUN-1998; 98US-0090254.
PR 23-JUN-1998; 98US-0090349.
PR 23-JUN-1998; 98US-0090355.
PR 24-JUN-1998; 98US-0090429.
PR 24-JUN-1998; 98US-0090431.
PR 24-JUN-1998; 98US-0090435.
PR 24-JUN-1998; 98US-0090444.
PR 24-JUN-1998; 98US-0090445.
PR 24-JUN-1998; 98US-0090461.
PR 24-JUN-1998; 98US-0090472.
PR 24-JUN-1998; 98US-0090535.
PR 24-JUN-1998; 98US-0090538.
PR 24-JUN-1998; 98US-0090540.
PR 24-JUN-1998; 98US-0090557.
PR 25-JUN-1998; 98US-0090676.
PR 25-JUN-1998; 98US-0090678.
PR 25-JUN-1998; 98US-0090688.
PR 25-JUN-1998; 98US-0090690.
PR 25-JUN-1998; 98US-0090691.
PR 25-JUN-1998; 98US-0090694.
PR 25-JUN-1998; 98US-0090695.
PR 25-JUN-1998; 98US-0090696.
PR 26-JUN-1998; 98US-0090862.
PR 26-JUN-1998; 98US-0090863.
PR 01-JUL-1998; 98US-0091358.

PR 01-JUL-1998; 98US-0091360.
PR 01-JUL-1998; 98US-0091544.
PR 02-JUL-1998; 98US-0091478.
PR 02-JUL-1998; 98US-0091486.
PR 02-JUL-1998; 98US-0091519.
PR 02-JUL-1998; 98US-0091626.
PR 02-JUL-1998; 98US-0091628.
PR 02-JUL-1998; 98US-0091633.
PR 02-JUL-1998; 98US-0091646.
PR 02-JUL-1998; 98US-0091673.
PR 07-JUL-1998; 98US-0091978.
PR 07-JUL-1998; 98US-0091982.
PR 09-JUL-1998; 98US-0092182.
PR 10-JUL-1998; 98US-0092472.
PR 20-JUL-1998; 98US-0093339.
PR 30-JUL-1998; 98US-0094651.
PR 04-AUG-1998; 98US-0095282.
PR 04-AUG-1998; 98US-0095285.
PR 04-AUG-1998; 98US-0095301.
PR 04-AUG-1998; 98US-0095302.
PR 04-AUG-1998; 98US-0095318.
PR 04-AUG-1998; 98US-0095321.
PR 04-AUG-1998; 98US-0095325.
PR 10-AUG-1998; 98US-0095916.
PR 10-AUG-1998; 98US-0095929.
PR 10-AUG-1998; 98US-0096012.
PR 11-AUG-1998; 98US-0096143.
PR 11-AUG-1998; 98US-0096146.
PR 12-AUG-1998; 98US-0096329.
PR 17-AUG-1998; 98US-0096757.
PR 17-AUG-1998; 98US-0096766.
PR 17-AUG-1998; 98US-0096768.
PR 17-AUG-1998; 98US-0096773.
PR 17-AUG-1998; 98US-0096791.
PR 17-AUG-1998; 98US-0096867.
PR 17-AUG-1998; 98US-0096891.
PR 17-AUG-1998; 98US-0096894.
PR 17-AUG-1998; 98US-0096895.
PR 17-AUG-1998; 98US-0096897.
PR 18-AUG-1998; 98US-0096949.
PR 18-AUG-1998; 98US-0096950.
PR 18-AUG-1998; 98US-0096959.
PR 18-AUG-1998; 98US-0096960.
PR 18-AUG-1998; 98US-0097022.
PR 19-AUG-1998; 98US-0097141.
PR 20-AUG-1998; 98US-0097218.
PR 24-AUG-1998; 98US-0097661.
PR 26-AUG-1998; 98US-0097951.
PR 26-AUG-1998; 98US-0097952.
PR 26-AUG-1998; 98US-0097954.
PR 26-AUG-1998; 98US-0097955.
PR 26-AUG-1998; 98US-0097971.
PR 26-AUG-1998; 98US-0097974.
PR 26-AUG-1998; 98US-0097978.
PR 26-AUG-1998; 98US-0097979.
PR 26-AUG-1998; 98US-0097986.
PR 26-AUG-1998; 98US-0098014.
PR 31-AUG-1998; 98US-0098525.
PR 16-SEP-1998; 98US-0100634.
PR 12-JAN-1999; 99US-0115565.

XX
PA (GETH) GENENTECH INC.
XX
PI Baker K, Chen J, Goddard A, Gurney AL, Smith V, Watanabe CK;
PI Wood WI, Yuan J;
XX
XX WPI: 2000-072883/06.
DR P-PSDB; AY66749.
XX
XX Membrane-bound proteins and related nucleotide sequences -
PT
XX
PS Claim 2; Fig 273; 822pp; English.
XX

CC AAH82377 to AAH83878 represent human ovarian tumour-associated
CC polynucleotide sequences which encode ovarian tumour proteins. The
CC ovarian tumour protein and polynucleotide sequences have cytostatic
CC activity, and can be used in gene therapy and vaccine production. The
CC ovarian tumour proteins and polynucleotides can be used to inhibit
CC the development of cancer, particularly ovarian cancer. They can also
CC be used to diagnose the onset and progression of cancer.

XX
SQ Sequence 476 BP; 142 A; 97 C; 110 G; 126 T; 1 other;

Query Match 47.6%; Score 51.4; DB 22; Length 476;
Best Local Similarity 96.9%; Pred. No. 1.8;
Matches 63; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

OY 6 ttgtatcatgtgagatcttaataaa-tgtgaaaaaatgaataaagcatgtttggtt 64
DB 178 TTTGTATATGTGAGATGTTTAAATTAATTGTGAAAAAATGAATTAAGCATGTTTGT 119
OY 65 tttcca 69
DB 118 TTCCA 114

RESULT 12

AA56243
ID AAX56243 standard; cDNA; 2802 BP.

XX
AC AAX56243;

XX 16-JUL-1999 (first entry)

XX Human vitamin D receptor related gamma 2 protein encoding cDNA.

DE
XX
KW Human: vitamin D receptor related protein; VDRR: obesity; diabetes;
KW anorexia; rheumatoid arthritis; lipoprotein defect; hyperlipidaemia;
KW hypercholesterolaemia; hyperlipoproteinaemia; osteoporosis; tumour;
KW hyperproliferative skin disorder; hyperthyroidism; ss.

OS Homo sapiens.

XX
PN WO919354-A1.

XX 22-APR-1999.

XX 31-AUG-1998; 98WO-SE01548.

XX
PR 31-MAR-1998; 98SE-0001148.

XX 14-OCT-1997; 97SE-0003745.

XX (PHAA) PHARMACIA & UPJOHN AB.

XX Berkenstam A, Dahlberg M;

XX WPI: 1999-302508/25.

XX P-PSDB: AAY09516.

XX New vitamin D receptor related (VDRR) polypeptides, useful for

XX treating obesity, diabetes, anorexia and rheumatoid arthritis

XX Claim 2; Page 21-22; 35pp; English.

XX The present sequence encodes a human vitamin D receptor related (VDRR)
XX polypeptide. Human VDRR polypeptides and substances which affect VDRR
XX signal transduction, can be used for treating metabolic, proliferative
XX or inflammatory conditions. They can be used in the manufacture of a
XX medicament for treating the following conditions: obesity, diabetes,
XX anorexia, lipoprotein defects, hyperlipidaemia, hypercholesterolaemia or
XX hyperlipoproteinaemia and osteoporosis, rheumatoid arthritis, benign and
XX malign tumours, hyperproliferative skin disorders or hyperthyroidism.
XX Nucleic acid vectors encoding for expression of a VDRR polypeptide can
XX be used for treating metabolic, proliferative or inflammatory conditions,
XX by introducing them into a mammal. The introduced nucleic acid is then

CC capable of transforming a cell in vivo and then polypeptide is expressed.
CC A substance affecting VDRR signal transduction, such as an agonist or
CC antagonist can be used for the manufacture of a medicament for treating
CC metabolic, proliferative or inflammatory condition.
CC N.B. The specification specifically claims the VDRR nucleic acid and
CC polypeptide sequences given in figures 1, 4, 7 and 8, but no figures
CC are given in the specification.

XX
SQ Sequence 2802 BP; 723 A; 715 C; 755 G; 609 T; 0 other;

Query Match 47.4%; Score 51.2; DB 20; Length 2802;
Best Local Similarity 77.5%; Pred. No. 1.4;
Matches 62; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

OY 24 ttaataaatgtgaaaaaatgaataaagcatgtttgtttccaaagaagaaaaaaa 83
DB 2714 tcaaatcaaggcaaaagaaatgaatgaatgtactttgtcctaaaaaagaaaaaaa 2773
OY 84 aaaaaaaaaaaaaaaaaa 103
DB 2774 aaaaaaaaaaaaaaaaaa 2793

RESULT 13

AA56242
ID AAX56242 standard; cDNA; 2910 BP.

XX
AC AAX56242;

XX 16-JUL-1999 (first entry)

XX Human vitamin D receptor related gamma protein encoding cDNA.

DE
XX
KW Human: vitamin D receptor related protein; VDRR: obesity; diabetes;
KW anorexia; rheumatoid arthritis; lipoprotein defect; hyperlipidaemia;
KW hypercholesterolaemia; hyperlipoproteinaemia; osteoporosis; tumour;
KW hyperproliferative skin disorder; hyperthyroidism; ss.

OS Homo sapiens.

XX
PN WO919354-A1.

XX 22-APR-1999.

XX 31-AUG-1998; 98WO-SE01548.

XX
PR 31-MAR-1998; 98SE-0001148.

XX 14-OCT-1997; 97SE-0003745.

XX (PHAA) PHARMACIA & UPJOHN AB.

XX Berkenstam A, Dahlberg M;

XX WPI: 1999-302508/25.

XX P-PSDB: AAY09515.

XX New vitamin D receptor related (VDRR) polypeptides, useful for

XX treating obesity, diabetes, anorexia and rheumatoid arthritis

XX Claim 2; Page 17-18; 35pp; English.

XX The present sequence encodes a human vitamin D receptor related (VDRR)
XX polypeptide. Human VDRR polypeptides and substances which affect VDRR
XX signal transduction, can be used for treating metabolic, proliferative
XX or inflammatory conditions. They can be used in the manufacture of a
XX medicament for treating the following conditions: obesity, diabetes,
XX anorexia, lipoprotein defects, hyperlipidaemia, hypercholesterolaemia or
XX hyperlipoproteinaemia and osteoporosis, rheumatoid arthritis, benign and
XX malign tumours, hyperproliferative skin disorders or hyperthyroidism.
XX Nucleic acid vectors encoding for expression of a VDRR polypeptide can
XX be used for treating metabolic, proliferative or inflammatory conditions,
XX by introducing them into a mammal. The introduced nucleic acid is then

CC	capable of transforming a cell in vivo and then polypeptide is expressed.
CC	A substance affecting VDR signal transduction, such as an agonist or
CC	antagonist can be used for the manufacture of a medicament for treating
CC	metabolic, proliferative or inflammatory condition.
CC	N.B. The specification specifically claims the VDR nucleic acid and
CC	polypeptide sequences given in figures 1, 4, 7 and 8, but no figures
CC	are given in the specification.
XX	
XX	
SQ	Sequence 2910 BP; 765 A; 728 C; 780 G; 636 T; 1 other;
XX	
Query Match	47.4%; Score 51.2; DB 20; Length 2910;
Best Local Similarity	77.5%; Pred. No. 1.4;
Matches	62; Conservative 0; Mismatches 18; Indels 0; Gaps 0
OY	24 ttaataaatgctgtaaaaaatgaataaagcattgttgcctccaaagaaaaaaa 83
DB	2818 tcaatcaaaagcacaaggaattaaataatgctactcttgctctcaaaaaa 2877
OY	84 aaaaaaaaaaaaaaaaaa 103
DB	2878 aaaaaaaaaaaaaaaaaa 2897
RESULT	14
AAA26358	
ID	AAA26358 standard; cDNA; 1052 BP.
XX	
AC	AAA26358;
XX	
DT	29-JUN-2000 (first entry)
XX	
DE	Human secreted protein gene 13 SEQ ID NO:23.
XX	
KW	Human; secreted protein; diagnosis; cytostatic; immunosuppressive;
KW	antiHIV; antiInflammatory; nootropic; neuroprotective; antiAllergic;
KW	osteoporatic; antiarthritic; antibacterial; antidiabetic; antiasthma;
KW	antiporiatic; cardiant; gene therapy; cancer; neurological disorder;
KW	immune disease; inflammation; blood disorder; tumour; chromosome 16; ss.
XX	
OS	Homo sapiens.
XX	
PN	WO200006698-A1.
XX	
PD	10-FEB-2000.
XX	
PF	29-JUL-1999; 99MO-US17130.
XX	
PR	30-JUL-1998; 98US-0094657.
PR	05-AUG-1998; 98US-0095486.
PR	06-AUG-1998; 98US-0095454.
PR	06-AUG-1998; 98US-0095455.
PR	12-AUG-1998; 98US-0096319.
XX	
PA	(HUMA-) HUMAN GENOME SCI INC.
XX	
PI	Komatsoulis GA, Rosen CA, Ruben SM, Duan R, Moore PA, Shi Y;
PI	Lafleur D, Wei Y, Ni J, Florence KA, Young PE, Brewer LA;
PI	Soppet DR, Endress GA, Edner R, Olsen HS, Mucenski M;
XX	
XX	
DR	WPI: 2000-195282/17.
XX	
XX	P-PSDB: AAY91463.
PT	New isolated human genes and the secreted polypeptides they encode,
PT	useful for diagnosis and treatment of e.g. cancers, neurological
PT	disorders, immune diseases, inflammation or blood disorders -
XX	
XX	
PS	Claim 1; Page 380-381; 634pp; English.
XX	
XX	
CC	The polynucleotide sequences given in AAA26346 to AAA26458 encode the
CC	human secreted proteins given in AAY91451 to AAY91691. The human secreted
CC	proteins can have activities based on the tissues and cells they are
CC	expressed in. Examples of the activities are: cytostatic;

Query Match	Best Local Similarity	Score 51;	DB 21;	Length 1052;		
Matches 57;	Conservative	85.1%;	Pred No. 1.7;	Mismatches 10;	Indels 0;	Gaps 0
QY 37	aaaaaaaaaataagcatgttctggtttccaaaagaaaaaaaaaaaaaaaaa	96				
DB 936	aaaaaaaaataaaaaaaatttttaatttttaaaaaaaaaaaaaaaaaa	995				
QY 97	aaaaaaa 103					
DB 996	aaaaaaa 1002					
<p>RESULT 15</p> <p>AAZ36918</p> <p>ID AAZ36918 standard; DNA; 1064 BP.</p> <p>XX AC AAZ36918;</p> <p>XX DT 13-MAR-2000 (first entry)</p> <p>XX DE</p> <p>XX DNA encoding Xenopus Geminin-H protein.</p> <p>XX DE</p> <p>XX Nuclear protein; Geminin-L; DNA replication; neurogenesis; neural growth;</p> <p>XX Geminin-H; isoform; pre-replication complex; origin of replication;</p> <p>XX nuclear DNA replication; anaphase promoting complex; ectodermal cell;</p> <p>XX nuclear cell; neural hypertrophy; ectopic neurogenesis;</p> <p>XX future neural plate; early gastrula; bone morphogenetic protein 4; BMP4;</p> <p>XX proliferative disease; cancer; neurological disease; Parkinson's disease;</p> <p>XX Alzheimer's disease; multiple sclerosis; spinal cord injury; ss.</p> <p>XX</p> <p>OS Xenopus sp.</p> <p>XX</p> <p>XX Key Location/Qualifiers</p> <p>XX CDS 53..712</p> <p>XX FT /tag= a</p> <p>XX FT /product= "Geminin-H"</p> <p>XX</p> <p>XX WO9958673-A1.</p> <p>XX PN</p> <p>XX 18-NOV-1999.</p> <p>XX PD</p> <p>XX 13-MAY-1999; 99WO-US10577.</p> <p>XX PE</p> <p>XX 13-MAY-1998; 98US-0085371.</p> <p>XX PR 11-JUN-1998; 98US-0096724.</p> <p>XX</p>						

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 17, 2002, 05:58:21 ; Search time 93.06 Seconds

(without alignments)
262.837 Million cell updates/sec

Title: US-09-823-101-6

Perfect score: 108
Sequence: 1 acctatttgatctgtgagc.....aaaaaaaaaaagcgcg 108

Scoring table: IDENTITY_NUC
Gap 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Issued Patents.NA:*
1: /cgn2_6/ptodata/2/lna/5A.COMB.seq:*
2: /cgn2_6/ptodata/2/lna/5B.COMB.seq:*
3: /cgn2_6/ptodata/2/lna/6A.COMB.seq:*
4: /cgn2_6/ptodata/2/lna/6B.COMB.seq:*
5: /cgn2_6/ptodata/2/lna/PCtUS.COMB.seq:*
6: /cgn2_6/ptodata/2/lna/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	61.8	57.2	982	4	US-08-943-731-110 Sequence 110, Appl
2	61.8	57.2	38682	4	US-08-943-731-2 Sequence 2, Appl
3	48.4	44.8	3138	1	US-07-867-106-4 Sequence 4, Appl
4	48	44.4	1604	1	US-08-665-966-9 Sequence 9, Appl
5	48	44.4	1604	3	US-09-041-780-9 Sequence 9, Appl
6	47.8	44.3	2255	4	US-08-871-572B-3 Sequence 3, Appl
7	47.2	43.7	1558	1	US-08-455-550-7 Sequence 7, Appl
8	47.2	43.7	5613	2	US-08-463-418-1 Sequence 1, Appl
9	46.8	43.3	3637	1	US-08-445-640-3 Sequence 3, Appl
10	46.8	43.3	3637	3	US-08-170-558-3 Sequence 3, Appl
11	46.8	43.3	3637	3	US-08-447-314-3 Sequence 3, Appl
12	46.8	43.3	3637	3	US-08-445-461-3 Sequence 3, Appl
13	46.8	43.3	4931	4	US-08-726-320-2 Sequence 2, Appl
14	46.8	43.3	4931	4	US-09-208-716-2 Sequence 2, Appl
15	46.8	43.3	5852	1	US-07-867-106-2 Sequence 2, Appl
16	46.6	43.1	688	6	5498694-3 Patent No. 5498694
17	46.4	43.0	127	3	US-09-014-416-59 Sequence 59, Appl
18	46.4	43.0	126	3	US-09-014-416-63 Sequence 63, Appl
19	46.4	43.0	183	3	US-09-014-416-60 Sequence 60, Appl
20	46.4	43.0	200	3	US-09-014-416-64 Sequence 64, Appl
21	46.4	43.0	9595	3	US-09-014-416-4 Sequence 4, Appl
22	46.4	43.0	9599	3	US-09-014-416-2 Sequence 2, Appl
23	46.4	43.0	9599	3	US-09-014-416-6 Sequence 6, Appl
24	46.2	42.8	742	1	US-07-847-010-12 Sequence 12, Appl
25	45.8	42.4	198	1	US-08-330-108-16 Sequence 16, Appl
26	45.8	42.4	198	5	PCR-US92-10087-16 Sequence 16, Appl
27	45.8	42.4	1332	2	US-09-057-762-1 Sequence 1, Appl

28	45.8	42.4	1332	3	US-08-326-119A-1 Sequence 1, Appl
29	45.8	42.4	3581	2	US-08-738-349-1 Sequence 1, Appl
30	45.6	42.2	1582	3	US-08-545-196B-10 Sequence 10, Appl
31	45.6	42.2	1582	3	US-08-545-196B-12 Sequence 12, Appl
32	45.4	42.0	1117	4	US-09-247-373B-33 Sequence 33, Appl
33	45.4	42.0	1559	2	US-08-417-174-1 Sequence 1, Appl
34	45.4	42.0	1559	2	US-08-231-565A-1 Sequence 1, Appl
35	45.4	42.0	1559	2	US-09-007-961-1 Sequence 1, Appl
36	45.4	42.0	1559	4	US-09-267-439-1 Sequence 1, Appl
37	45.2	41.9	1441	4	US-08-821-994-63 Sequence 63, Appl
38	45.2	41.9	1813	4	US-09-071-224-3 Sequence 3, Appl
39	45.2	41.9	3410	4	US-09-020-956-110 Sequence 110, App
40	45.2	41.9	3410	4	US-09-030-607-110 Sequence 110, App
41	45.2	41.9	3572	2	US-08-713-815A-2 Sequence 2, Appl
42	45	41.7	780	2	US-08-540-804-36 Sequence 36, Appl
43	45	41.7	780	4	US-08-590-399-36 Sequence 36, Appl
44	45	41.7	1639	4	US-09-362-473-5 Sequence 5, Appl
45	45	41.7	3848	4	US-09-112-096-28 Sequence 28, Appl

ALIGNMENTS

RESULT 1
US-08-943-731-110
Sequence 110, Application US/08943731
Patent No. 6265157
GENERAL INFORMATION:
APPLICANT: PROCKOP, DARWIN J.
APPLICANT: SPOTILA, LORETTA D.
APPLICANT: DELTAS, CONSTANTINOS D.
APPLICANT: SEREDA, LARISSA
APPLICANT: LARSON, ANDREA W.
APPLICANT: PACK, MICHAEL
APPLICANT: COLIGE, ALAIN
APPLICANT: EARLY, JAMES
APPLICANT: KORRKO, JARMO
APPLICANT: ALA-KOKKO, LEENA, et al.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DETECTING
TITLE OF INVENTION: ALTERED TYPE I OR TYPE IX COLLAGEN GENE SEQUENCES
NUMBER OF SEQUENCES: 666
CORRESPONDENCE ADDRESS:
ADDRESSEE: PANTICH SCHWARZE JACOBS & NADEL, P.C.
STREET: ONE COMMERCE SQUARE, 2005 MARKET STREET, 22ND
STREET: FLR.
CITY: PHILADELPHIA
STATE: PA
COUNTRY: USA
ZIP: 19103-7086
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/943,731
FILING DATE: 03-OCT-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/212,322
FILING DATE: 14-MAR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/803,628
FILING DATE: 03-DEC-1991
ATTORNEY/AGENT INFORMATION:
NAME: DOYLE LEARY Ph.D., KATHRYN
REGISTRATION NUMBER: 36,317
REFERENCE/DOCKET INFORMATION: 9598-27
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-965-1284
TELEFAX: 215-567-2991
TELEX: 831-494
INFORMATION FOR SEQ ID NO: 110:

SEQUENCE CHARACTERISTICS:
LENGTH: 982 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-943-731-110

Query Match 57.2%; Score 61.8; DB 4; Length 982;
Best Local Similarity 90.6%; Pred. No. 0.00037;
Matches 77; Conservative 0; Mismatches 7; Indels 1; Gaps 1;

QY 6 ttgtatattgtgagatgtttaataaa-tgtgaaaaaaatgaataaagcatgtttggt 64
|||||
Db 771 TTTGTATATGTGAGATGTTTAAATAATTGTGAATAAATGAATGAATGAATGTTGTT 830

QY 65 ttccaaagaaaaaataaaaaaa 89
|||||
Db 831 TTCCAAAGAACATATTGAGTAA 855

RESULT 2

US-08-943-731-2
Sequence 2, Application US/08943731

Patent No. 6265157

GENERAL INFORMATION:

APPLICANT: PROCKOP, DARWIN J.

APPLICANT: SPOTILA, LORETTA D.

APPLICANT: DELTAS, CONSTANTINOS D.

APPLICANT: SEREDA, LARISA

APPLICANT: LARSON, ANDREA W.

APPLICANT: PACK, MICHAEL

APPLICANT: COLIGE, ALAIN

APPLICANT: EARLY, JAMES

APPLICANT: KORRKO, JARMO

APPLICANT: ALA-KORRKO, LERNA, et al.

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DETECTING

NUMBER OF SEQUENCES: 666

CORRESPONDENCE ADDRESS:

ADDRESSEE: PANITCH SCHWARZ JACOBS & NADEL, P.C.

STREET: ONE COMMERCE SQUARE, 2005 MARKET STREET, 22ND

STREET: FLR.

CITY: PHILADELPHIA

STATE: PA

COUNTRY: USA

ZIP: 19103-7086

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/943,731

FILING DATE: 03-OCT-1997

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/212,322

FILING DATE: 14-MAR-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/803,628

FILING DATE: 03-DEC-1991

ATTORNEY/AGENT INFORMATION:

NAME: DOYLE LEARY Ph.D., KATHRYN

REGISTRATION NUMBER: 36,317

REFERENCE/DOCKET NUMBER: 9598-27

TELEPHONE: 215-965-1284

TELEPHONE: 215-567-2991

TELEX: 831-494

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 38682 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-943-731-2

Query Match 57.2%; Score 61.8; DB 4; Length 38682;
Best Local Similarity 90.6%; Pred. No. 0.00027;
Matches 77; Conservative 0; Mismatches 7; Indels 1; Gaps 1;

QY 6 ttgtatattgtgagatgtttaataaa-tgtgaaaaaaatgaataaagcatgtttggt 64
|||||
Db 38471 TTTGTATATGTGAGATGTTTAAATAATTGTGAATAAATGAATGAATGTTGTT 38530

QY 65 ttccaaagaaaaaataaaaaaa 89
|||||
Db 38531 TTCCAAAGAACATATTGAGTAA 38555

RESULT 3

US-07-867-106-4/C
Sequence 4, Application US/07867106

Patent No. 5389526

GENERAL INFORMATION:

APPLICANT: SLADE, Martin B

APPLICANT: Chang, Andy C M

APPLICANT: Williams, Keith L

TITLE OF INVENTION: Improved plasmid Vectors for Cellular

NUMBER OF SEQUENCES: 19

CORRESPONDENCE ADDRESS:

ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5389526rfs

STREET: One Liberty Place 46th Floor

CITY: Philadelphia

STATE: PA

COUNTRY: USA

ZIP: 19103

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/867,106

FILING DATE: 19920625

PRIOR APPLICATION DATA:

APPLICATION NUMBER: AU PJ 7187

APPLICATION NUMBER: PCT/AU90/00530

FILING DATE: 02-NOV-1989

ATTORNEY/AGENT INFORMATION:

NAME: Feeney, Joanne Longo

REGISTRATION NUMBER: 35,134

REFERENCE/DOCKET NUMBER: RICE-0002

TELECOMMUNICATION INFORMATION:

TELEPHONE: 215-568-3439

TELEPHONE: 215-568-3100

TELEPHONE: 215-568-3439

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 3138 base pairs

TYPE: NUCLEIC ACID

STRANDEDNESS: single

TOPOLOGY: circular

MOLECULE TYPE: DNA (genomic)

ANTI-SENSE: NO

US-07-867-106-4

Query Match 44.8%; Score 48.4; DB 1; Length 3138;

Best Local Similarity 68.4%; Pred. No. 0.084;

Matches 67; Conservative 0; Mismatches 31; Indels 0; Gaps 0;


```

? APPLICATION NUMBER: 08/056,564
? FILING DATE: 30-APR-1993
? APPLICATION NUMBER: 07/577,892
? FILING DATE: 05-SEP-1990
? ATTORNEY/AGENT INFORMATION:
? NAME: Eisenstein, Ronald I
? REGISTRATION NUMBER: 30628
? REFERENCE/DOCKET NUMBER: 40302-FWC-DIV
TELECOMMUNICATION INFORMATION:
? TELEPHONE: 617-523-3400
? TELEFAX: 617-523-6440
? TELEX: 200291
? INFORMATION FOR SEQ ID NO: 7:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 1558 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: cDNA
? HYPOTHEetical: NO
? ANTI-SENSE: NO
? FRAGMENT TYPE:
? ORIGINAL SOURCE:
US-08-455-550-7

Query Match          .43.7%   Score 47.2; DB 1; Length 1558;
Best Local Similarity 72.6%; Pred. No. 0.15; Indels    0; Gaps    0;
Matches      61; Conservative    0; Mismatches     23;

```

```

Query Match 43.7%; Score 47.2; DB 1; Length 1558;
Best Local Similarity 72.6%; Pred. No. 0.15;
Matches 61; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

Oy 20 atgttaataaattgaaaaaataagcaatglttggtttccaaagaaaaa 79
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1450 ACGTATTAAAGAAATAAAATAAATTAAGTCATTATTTTAAACATAAAAA 1509
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Oy 80 aaaaaaaaaaaaaaaaaaaaaa 103
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1510 AAAAAAAAAAAAAAAAAAAAAA 1533
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 8
US-08-463-418-1/c
Sequence 1, Application US/08463418
Patent No. 5908971
GENERAL INFORMATION:
  APPLICANT: Van Der Straeten, Dominique et al.
  TITLE OF INVENTION: CRUCIFER ACC SYNTHASE AND USES THEREOF
  NUMBER OF SEQUENCES: 10
  CORRESPONDENCE ADDRESS:
  ADDRESSEE: Fish & Richardson P.C.
  STREET: 225 Franklin Street
  CITY: Boston
  STATE: MA
  COUNTRY: USA
  ZIP: 02110-2804
COMPUTER READABLE FORM:
  MEDIUM TYPE: Floppy disk
  COMPUTER: IBM PC compatible
  OPERATING SYSTEM: PC-DOS/MS-DOS
  SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
  APPLICATION NUMBER: US/08/463,418
  FILING DATE: 05-JUN-1995
  CLASSIFICATION: 800
PRIOR APPLICATION DATA:
  APPLICATION NUMBER: US 07/962,481
  FILING DATE: 15-OCT-1992
ATTORNEY/AGENT INFORMATION:
  NAME: Clark, Paul T.
  REGISTRATION NUMBER: 30,162
  REFERENCE/DOCKET NUMBER: 00786/161002
TELECOMMUNICATION INFORMATION:
  TELEPHONE: 617/542-5070
  TELEFAX: 617/542-8906

```

```
TELEX: 200154
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5613 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-463-418-1

Query Match      43.7%; Score 47.2; DB 2; Length 5613;
Best Local Similarity 72.1%; Pred. No. 0.13;
Matches 75; Conservative 0; Mismatches 28; Indels 1; Gaps 1;

QY 5 attgtatcgtgagatggttaataatgtaaaaaatgaataagcatgttgcgtt 64
    ||||| ||| | | | | | | | | | | | | | | | | | | | | | | |
Db 2148 ATTATATTAGGCTAAGTTAATA-AAAATAAGTACTAATTAATGCAACGAACTTT 2090

QY 65 ttccaaaagaaaaaagaaaaaagaaaaaagcgcgcg 108
    | |||| | | | | | | | | | | | | | | | | | | | | | | | |
Db 2089 AACAAAAAAGAAAAAAGAAAAAAGAAAAAAGAACACG 2046

RESULT 9
US-08-445-640-3
; Sequence 3, Application US/08445640
; Patent No. 5709858
; GENERAL INFORMATION:
; APPLICANT: Godowski, Paul J.
; APPLICANT: Mark, Melanie R.
; APPLICANT: Scadden, David T.
; APPLICANT: Baker, Kevin P.
; APPLICANT: Baron, Will F.
; TITLE OF INVENTION: Protein Tyrosine Kinases
; NUMBER OF SEQUENCES: 35
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 kb floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/445,640
; FILING DATE: 22-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/170558
; FILING DATE: 20-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/157563
; FILING DATE: 23-NOV-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Hasak, Janet E.
; REGISTRATION NUMBER: 28,616
; REFERENCE/DOCKET NUMBER: 854C2
; TELEPHONE: 415/225-1896
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3637 bases
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-445-640-3
```

```
Query Match      43.3%; Score 46.8; DB 1; Length 3637;
Best Local Similarity 70.0%; Pred. No. 0.16;
Matches 63; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

QY 14 tctgagatggttaataatgtaaaaaatgaataagcatgttgcgttccaaaag 73
    ||| | | | | | | | | | | | | | | | | | | | | | | | | |
Db 3539 TGAGCAATTTTATCCCTCAGTACGACGATATAATAAGTTGAGTTTCACAAA 3598

QY 74 aaaaaaagaaaaaagaaaaaagaaaaaagcgcgcg 103
    ||||| | | | | | | | | | | | | | | | | | | | | | | |
Db 3599 AAAAAAAGAAAAAAGAAAAAAGAAAAAAGAACACG 3628

RESULT 10
US-08-170-558-3
; Sequence 3, Application US/08170558
; Patent No. 6001621
; GENERAL INFORMATION:
; APPLICANT: Godowski, Paul J.
; APPLICANT: Mark, Melanie R.
; APPLICANT: Scadden, David T.
; APPLICANT: Baker, Kevin P.
; APPLICANT: Baron, Will F.
; TITLE OF INVENTION: Protein Tyrosine Kinases
; NUMBER OF SEQUENCES: 35
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 kb floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/170,558
; FILING DATE: 20-DEC-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/157563
; FILING DATE: 23-NOV-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Hasak, Janet E.
; REGISTRATION NUMBER: 28,616
; REFERENCE/DOCKET NUMBER: 854C1
; TELEPHONE: 415/225-1896
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3637 bases
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-170-558-3

Query Match      43.3%; Score 46.8; DB 3; Length 3637;
Best Local Similarity 70.0%; Pred. No. 0.16;
Matches 63; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

QY 14 tctgagatggttaataatgtaaaaaatgaataagcatgttgcgttccaaaag 73
    ||| | | | | | | | | | | | | | | | | | | | | | | | | |
Db 3539 TGAGCAATTTTATCCCTCAGTACGACGATATAATAAGTTGAGTTTCACAAA 3598

QY 74 aaaaaaagaaaaaagaaaaaagaaaaaagcgcgcg 103
    ||||| | | | | | | | | | | | | | | | | | | | | | | |
```

Db 3599 AAAAAAAAAAAAAAAAAAAAAAAAAA 3628

RESULT 11

US-08-447-314-3
; Sequence 3, Application US/08447314
; Patent No. 6087144

GENERAL INFORMATION:

APPLICANT: Scadden, David T.
APPLICANT: Baker, Kevin P.
APPLICANT: Baron, Will F.
TITLE OF INVENTION: Protein Tyrosine Kinases
NUMBER OF SEQUENCES: 35
CORRESPONDENCE ADDRESS:

ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080

COMPUTER READABLE FORM:

MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: patin (Genentech)

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/447,314
FILING DATE: 22-MAY-1995

CLASSIFICATION: 435

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/170558
FILING DATE: 20-DEC-1993

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/157563

FILING DATE: 23-NOV-1993
ATTORNEY/AGENT INFORMATION:
NAME: Hasak, Janet E.

REGISTRATION NUMBER: 28,616
REFERENCE/DOCKET NUMBER: 854C1D2
TELECOMMUNICATION INFORMATION:

TELEPHONE: 415/225-1896
TELEFAX: 415/952-9881
TELEX: 910/371-7168

INFORMATION FOR SEQ. ID NO: 3:
SEQUENCE CHARACTERISTICS:

LENGTH: 3637 bases
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear

US-08-447-314-3

Query Match 43.3%; Score 46.8; DB 3; Length 3637;
Best Local Similarity 70.0%; Pred. No. 0.16; Mismatches 27; Indels 0; Gaps 0;

Matches 63; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

QY 14 tgtgagatgtttaataaataatgaataaagcatgttgcctcaaaag 73
||| | ||||| || | ||||| ||| ||||| ||| ||||| |||

Db 3539 TGAGGCAATTTTATCCCTGACACTAGGCGATTAATAAGGTGAGTTTCCACAAA 3598

QY 74 aaaaaaaaaaaaaaaaaaaaaaaaaa 103
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 3599 AAAAAAAAAAAAAAAAAAAAAAAAAA 3628

RESULT 12

US-08-445-461-3
; Sequence 3, Application US/08445461
; Patent No. 6096527

GENERAL INFORMATION:

APPLICANT: Godowski, Paul J.
APPLICANT: Mark, Melanie R.
APPLICANT: Scadden, David T.

APPLICANT: Baker, Kevin P.

APPLICANT: Baron, Will F.

TITLE OF INVENTION: Protein Tyrosine Kinases

NUMBER OF SEQUENCES: 35

CORRESPONDENCE ADDRESS:

ADDRESSEE: Genentech, Inc.

STREET: 460 Point San Bruno Blvd

CITY: South San Francisco

STATE: California

COUNTRY: USA

ZIP: 94080

COMPUTER READABLE FORM:

MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: patin (Genentech)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/445,461

FILING DATE: 22-MAY-1995

CLASSIFICATION: 530

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/170558

FILING DATE: 20-DEC-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/157563

FILING DATE: 23-NOV-1993

ATTORNEY/AGENT INFORMATION:

NAME: Hasak, Janet E.

REGISTRATION NUMBER: 28,616

REFERENCE/DOCKET NUMBER: 854C3

TELECOMMUNICATION INFORMATION:

TELEPHONE: 415/225-1896

TELEFAX: 415/952-9881

TELEX: 910/371-7168

INFORMATION FOR SEQ. ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 3637 bases

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

US-08-445-461-3

Query Match 43.3%; Score 46.8; DB 3; Length 3637;
Best Local Similarity 70.0%; Pred. No. 0.16; Mismatches 27; Indels 0; Gaps 0;

Matches 63; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

QY 14 tgtgagatgtttaataaataatgaataaagcatgttgcctcaaaag 73
||| | ||||| || | ||||| ||| ||||| ||| ||||| |||

Db 3539 TGAGGCAATTTTATCCCTGACACTAGGCGATTAATAAGGTGAGTTTCCACAAA 3598

QY 74 aaaaaaaaaaaaaaaaaaaaaaaaaa 103
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 3599 AAAAAAAAAAAAAAAAAAAAAAAAAA 3628

RESULT 13

US-08-726-320-2
; Sequence 2, Application US/08726320
; Patent No. 6171815

GENERAL INFORMATION:

APPLICANT: Au-Young, Janice

APPLICANT: Bandman, Olga

APPLICANT: Coleman, Roger

TITLE OF INVENTION: NOVEL HUMAN SULFONYLUREA RECEPTOR

NUMBER OF SEQUENCES: 5

CORRESPONDENCE ADDRESS:

ADDRESSEE: Incyte Pharmaceuticals, Inc.

STREET: 3174 Porter Drive

CITY: Palo Alto

STATE: CA

COUNTRY: U.S.

ZIP: 94304

```
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/726,320
FILING DATE: Filed Herewith
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: AF-0001 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-845-4166
TELEFAX: 415-845-0555
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 4931 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
IMMEDIATE SOURCE:
LIBRARY:
CLONE: Consensus
US-08-726-320-2

Query Match
Best Local Similarity 43.3%; Score 46.8; DB 4; Length 4931;
Matches 60; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

QY 22 gttaaataaatgtaaaataaagcatgttggtttccaaagaaaaaa 81
DB 4842 GGTTCCTACCTGAATACCTGTAAATGAATGATTGATTATTA 4901
QY 82 aaaaaaaaaaaaaaaaaa 103
DB 4902 AAAAAAAAAAAAAAAAAA 4923

RESULT 14
US-09-208-716-2
Sequence 2, Application US/09208716
Patent No. 6235880
GENERAL INFORMATION:
APPLICANT: Au-Young, Janice
APPLICANT: Bandman, Olga
APPLICANT: Coleman, Roger
TITLE OF INVENTION: NOVEL HUMAN SULFONYLUREA RECEPTOR
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto.
STATE: CA
COUNTRY: U.S.
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/208,716
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/726,320
FILING DATE:
ATTORNEY/AGENT INFORMATION:
```

```
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: AF-0001 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-845-0555
TELEFAX: 415-845-4166
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 4931 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
IMMEDIATE SOURCE:
LIBRARY:
CLONE: Consensus
US-09-208-716-2

Query Match
Best Local Similarity 43.3%; Score 46.8; DB 4; Length 4931;
Matches 60; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

QY 22 gttaaataaatgtaaaataaagcatgttggtttccaaagaaaaaa 81
DB 4842 GGTTCCTACCTGAATACCTGTAAATGAATGATTGATTATTA 4901
QY 82 aaaaaaaaaaaaaaaaaa 103
DB 4902 AAAAAAAAAAAAAAAAAA 4923

RESULT 15
US-07-867-106-2/c
Sequence 2, Application US/07867106
Patent No. 5389526
GENERAL INFORMATION:
APPLICANT: Slade, Martin B
APPLICANT: Chang, Andy C M
APPLICANT: Williams, Keith L
TITLE OF INVENTION: Improved Plasmid Vectors for Cellular
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & NO. 5389526r1s
STREET: One Liberty Place 46th Floor
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19103
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/867,106
FILING DATE: 19920625
PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU PJ 7187
APPLICATION NUMBER: PCT/AU90/00530
FILING DATE: 02-NOV-1989
ATTORNEY/AGENT INFORMATION:
NAME: Feeney, Joanne Longo
REGISTRATION NUMBER: 35,134
REFERENCE/DOCKET NUMBER: RICE-0002
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-568-3100
TELEFAX: 215-568-3439
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 5852 base pairs
TYPE: NUCLEIC ACID
```

```
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2378..5038
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2378..5038
; US-07-867-106-2
```

```
Query Match 43.3%; Score 46.8; DB 1; Length 5852;
Best Local Similarity 67.3%; Pred. No. 0.15;
Matches 66; Conservative 0; Mismatches 32; Indels 0; Gaps 0;
```

```
OY 6 ttgtatcgtgagatgttaataatgtaaaatgaataacatgcttgcgtt 65
||| ||| | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5697 TTTTATTATTATAATTATATATTAAATTAAATTAATTAATAAAAAA 5638

OY 66 tccaaagaaaaaataaaaaaataaaaaa 103
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5637 ATTTAAATTAAAAAATAAAAAAATAAAAAA 5600
```

Search completed: January 17, 2002, 05:58:24
Job time: 1032 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 17, 2002, 08:48:08 ; Search time 2206.96 Seconds
(without alignments)
4716.769 Million cell updates/sec

Title: US-09-823-101-5

Perfect score: 631
Sequence: 1 acaagctttttttttttt.....ccacagacaaaaaacacat 631

Scoring table:

IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: gb_da:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_om:*
20: em_or:*
21: em_ov:*
22: em_pat:*
23: em_ph:*
24: em_pl:*
25: em_ro:*
26: em_sts:*
27: em_sy:*
28: em_un:*
29: em_vi:*
30: em_htgo_hum:*
31: em_htgo_inv:*
32: em_htgo_rod:*
33: em_htg_hum:*
34: em_htg_inv:*
35: em_htg_rod:*
36: em_htg_other:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	87.8	13.9	396	6	AX093215	AX093215 Sequence
2	78.8	12.5	3747	9	HSU42766	U42766 Human neuro
3	78	12.4	1588	9	HSM800676	AL080159 Homo sapi
4	77	12.2	144	6	189947	Sequence 26
5	76.8	12.2	750	9	HSM802288	AL137550 Homo sapi
6	76.8	12.2	1555	9	AB063070	AB063070 Macaca fa
7	76.8	12.2	74896	2	AC093310	AC093310 Homo sapi
8	76.4	12.1	1999	9	HS242859	AJ242859 Homo sapi
9	76.2	12.1	1195	9	BC007326	BC007326 Homo sapi
10	76.2	12.1	4670	9	AF104032	AF104032 Homo sapi
11	76	12.0	396	6	AX093192	AX093192 Sequence
12	76	12.0	1483	9	AK026528	AK026528 Homo sapi
13	75.6	12.0	612	6	AX187054	AX187054 Sequence
14	75.6	12.0	1163	9	AK000614	AK000614 Homo sapi
15	75	11.9	580	9	HSM800144	AL050277 Homo sapi
16	74.8	11.9	1439	9	HSM802297	AL137557 Homo sapi
17	74.8	11.9	1765	9	HSM801934	AL137271 Homo sapi
18	74.6	11.8	1500	5	X1NKAATP	Y11587 X. laevis mr
19	74.6	11.8	3315	9	AB056768	AB056768 Macaca fa
20	74.4	11.8	1821	9	BC004951	BC004951 Homo sapi
21	74.4	11.8	2499	9	HSM801350	AL133016 Homo sapi
22	74.4	11.8	2755	10	AF158248	AF158248 Rattus no
23	74.4	11.8	4029	3	PLI249625	AJ249625 Paracentr
24	74.2	11.8	1983	10	BC005510	BC005510 Mus muscu
25	74.2	11.8	290452	2	AC079167	AC079167 Mus muscu
26	74	11.7	256	6	AX182154	AX182154 Sequence
27	74	11.7	559	9	HSM800249	AL049654 Homo sapi
28	74	11.7	1892	9	HSM801812	AL136844 Homo sapi
29	74	11.7	2463	10	BC005449	BC005449 Mus muscu
30	73.8	11.7	1422	9	AK026086	AK026086 Homo sapi
31	73.6	11.7	1347	10	BC008980	BC008980 Mus muscu
32	73.6	11.7	2336	8	AF353616	AF353616 Cronartiu
33	73.4	11.6	467	3	AF156168	AF156168 Butus ma
34	73.4	11.6	1283	10	AF017437	AF017437 Rattus no
35	73.4	11.6	1836	9	HSM802705	AL359596 Homo sapi
36	73.4	11.6	2090	9	BC004958	BC004958 Homo sapi
37	73.4	11.6	2253	10	BC010651	BC010651 Mus muscu
38	73.4	11.6	2417	9	HSM800967	AL117457 Homo sapi
39	73.4	11.6	2846	9	HSM802834	AL442082 Homo sapi
40	73.4	11.6	3513	9	HSM800944	AL117435 Homo sapi
41	73.4	11.6	6410	9	AF230496	AF230496 Homo sapi
42	73.2	11.6	447	9	AK026855	AK026855 Homo sapi
43	73.2	11.6	1414	10	BC005487	BC005487 Mus muscu
44	73.2	11.6	2097	9	BC001045	BC001045 Homo sapi
45	73.2	11.6	2271	9	AB056809	AB056809 Macaca fa

ALIGNMENTS

RESULT 1
AX093215
LOCUS AX093215 396 bp DNA
DEFINITION Sequence 33 from Patent WO0118046.
ACCESSION AX093215
VERSION AX093215.1 GI:13509664
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 396)
AUTHORS Xu,J. and Stolk,J.A.
TITLE Ovarian tumor sequences and methods of use therefor
JOURNAL Patent: WO 0118046-A 33 15-MAR-2001;
CORIXA CORPORATION (US)
FEATURES
source Location/Qualifiers
1..396
/organism="Homo sapiens"
/db_xref="taxon:9606"

[illegible]

COMMENT	FEATURES
Martinsried, GERMANY Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ). Email s.wiemann@dkfz-heidelberg.de; sequenced by MedGenomix (Martinsried/Germany) within the CDNA Sequencing consortium of the German Genome Project. This clone (DKFZ434C0820) is available at the RZPD in Berlin. Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 10559 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further Information about the clone and the sequencing project is available at http://www.mips.biochem.mpg.de/proj/cdna/ .	source 1..750 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="DKFZp434C0820" /clone_1ib="434 (synonym: htes3). Vector pSport1; host DH10B; sites NotI + SalI" /dev_stage="adult" /tissue_type="testis" 1..444 /gene="DKFZp434C0820" <1..444 /gene="DKFZp434C0820" /note="Similarity to cyclin B3 - chicken" /codon_start=1 /product="hypothetical protein" /protein_id="CA870806.1" /db_xref="GI:5808235" /translation="LVNKKDINPIAVHFLLRYARCIIHTNMKTLTSLRICEMTLQEE YHYOERASKIAASILLATYMKRLGVWPLETHYSKSYSELPILYROLKLITPSS YDSIRAYVYKISHPEVEVAKIRPLDMKLLEIINCCEAGGLV" 628..633
polyA_signal polyA_site BASE COUNT 265 a 147 c 126 g 212 t ORIGIN	
Query Match Best Local Similarity 12.2%; Score 76.8; DB 9; Length 750; Matches 90; Conservative 0; Mismatches 22; Indels 0; Gaps 0;	
QY 7 ttggggaaacctttc 66 Db 736 ttt 677 	
QY 67 tttttttttttttttttttttaaacacccccccttatatttttaagaca 118 Db 676 ttaaA 625 	
RESULT 6 AB063070/c LOCUS	
DEFINITION Macaca fascicularis brain cDNA clone:QJRA-10484, full insert sequence.	
ACCESSION AB063070	
VERSION AB063070.1 GI:14388527	
KEYWORDS oligo capping; fis (full insert sequence).	
SOURCE Macaca fascicularis adult male temporal lobe right cDNA to mRNA, clone_1lib:macaque brain cDNA library QJra clone:QJRA-10484. Macaca fascicularis Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea; Cercopithecinae; Macaca.	
ORGANISM	
REFERENCE AUTHORS Osada,N., Hida,M., Kusuda,J., Tanuma,R., Iseki,K., Hirai,M., Terao,K., Suzuki,Y., Sugano,S. and Hashimoto,K. Isolation of full-length cDNA clones from macaque brain cDNA libraries Unpublished 2 (bases 1 to 1555) Hashimoto,K., Osada,N., Hida,M., Kusuda,J. and Sugano,S. Direct Submission	

DEFINITION	Homo sapiens mRNA for langerin protein.
ACCESSION	AJ242859
VERSION	AJ242859.1 GI:5523278
KEYWORDS	langerin protein.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 1999) Valladeau,J., Ravel,O., Dezutter-Dambuyant,C., Moore,K., Kieljmeier,M., Duvert-Frances,V., Vincent,C., Schmitt,D., Davoust,J., Caux,C., Lebecque,S. and Sealand,S. langerin, a new transmembrane C-type lectin specific to Langerhans cells, induces the formation of Birbeck granules Unpublished 2 (bases 1 to 1999) Valladeau,J.M. Direct Submission Submitted (02-JUN-1999) Valladeau J.M., Laboratory for Immunological Research, Schering-Plough, 27 Chemin des Peupliers, 69571 Dardilly Cedex, FRANCE Revised by author 06-SEP-1999 Location/Qualifiers 1..1999 /organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="2" /cell_type="Langerhans cells" /map="2p13" 48..1034 /codon_start=-1 /product="langerin" /protein_id="CAB62403.1" /db_xref="GI:6523279"
JOURNAL	
AUTHORS	
TITLE	
REMARK	
FEATURES	
source	
CDS	
BASE COUNT	polyA-signal 1844 .. 1849 586 a 494 c 450 g 469 t
ORIGIN	
Query Match	12.1% Score 76.4 DB 9 Length 1999;
Best Local Similarity	59.0% Pred. No. 0.00023;
Matches 131 Conservative	0 Mismatches 91 Indels 0 Gaps 0;
OY 7	tcttttcttcttcttcttcttcttcttcttcttcttcttgsggaaacctttt 66
Dd 1973	TTTTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 1914
OY 67	tcttttcttcttcttcttcttcttctaacaaccccccttatattttagcagcgaccoccc 126
Dd 1913	TTTTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTGACCAGCA 1854
OY 127	acaaaaaggggggggcggtggaaaaacaactcaatgtgccttcgggggtcaaaccca 186
Dd 1853	GCAATTTTAATTGTGGAGAAGCTGGAAAGTGGAANTCAWGAAGGCACGAAAGCGCTGCCCA 1794
OY 187	ggagtagcttcttctgatgtaaaaaaaaaacagagaagagga 228
Dd 1793	TGCACAAGACACAGGAAGATGTGAAAATTCACAGAAGAGGAGA 1752
RESULT 9	
BC007326/c	1195 bp mRNA PRI 12-JUL-2001
LOCUS	BC007326 Homo sapIens, clone MGC:14879 IMAGE:3617525, mRNA, complete cds.
DEFINITION	BC007326
ACCESSION	BC007326.1 GI:13938381
VERSION	MGC.
KEYWORDS	

[illegible]

SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 4670)
AUTHORS	Prasad,P.D., Wang,H., Huang,W., Kekuda,R., Rajan,D.P., Leibach,F.H. and Ganapathy,V.
TITLE	Human LAT1, a subunit of system L amino acid transporter: molecular cloning and transport function
JOURNAL	Biochem. Biophys. Res. Commun. 255 (2), 283-288 (1999)
REFERENCE	2 (bases 1 to 4670) Prasad,P.D., Wang,H., Huang,W., Kekuda,R., Rajan,D.P., Leibach,F.H. and Ganapathy,V. Direct Submission Submitted (02-NOV-1998) Obstetrics and Gynecology, Medical College of Georgia, 1120 15th Street, Augusta, GA 30912, USA
FEATURES	Location/Qualifiers 1..4670 /organism="Homo sapiens" /db_xref="taxon:9606" /tissue_type="Placenta" 67..1590 /function="transports large neutral amino acids in association with 4F2HC antigen" /note="LAT1" /codon_start=1 /product="L-type amino acid transporter subunit LAT1" /protein_id="AAD20464.1" /db_xref="GI:4426640"
CDS	/translation="MACGPKRRRLAAPAEKEKREARKMLAKSADGSAPAGCEGV TLQKNLTNLNGLVAITVTIGSGIFVPPTGVLRKRGSLAVVMACGVPSIVGALD VALGGTITSKSGDYAYMLLEVYGSLPAPFLKLIMLLIRPSSOYIALVFATYLKP PPCTGPVEPAKKLVACLVCVLILTVNCNSGYAAARVODAFKAALKIALLILLGFVO IKGDVSNLDPNDFEFEGEKIDPVNIYALYGLFAGGMNYLNVENTEEMINPRNP AIIISLPVLTVLYVNLYAFETTLSTEDMSSEAVADEGNHLGWSMIIPRVGIS CEFGVNGLSFTSRPLFPVGRREGHLPIELIHOLTPRVSIVPGCVMTLLYPASRK IESFINPESPFNMCVALATTGMITWLRKRRELRPRIKYNIALVFPIIACLFIAYS FMKIPVECGIFFTTIISGLPYFFVGWMRNPKMWLOGISTIVLCOKLMOVPOET"
BASE COUNT	896 a 1471 c 1301 g 1002 t
ORIGIN	
Query Match:	12.1%; Score 76.2; Db 9; Length 4670;
Best Local Similarity	74.4%; Pred. No. 0.00025;
Matches	96; Conservative 0; Mismatches 33; Indels 0; Gaps 0;
Oy	7 ttttttttttttttttttttttttttttttttttcttg9gg9aaaacctttc 66 Db 4623 TTTTTTTT
Oy	67 tttttttttttttttttttttaaacaccocotttatlttagcacgacccccc 126 Db 4563 TTTTGTGGATGTGAAGTGAGTCCTCTTAATAAAAAATAAAACCOC 4504
Oy	127 acaaaaaag 135
Db	4503 ACAAANAAG 4495
RESULT	11
AX093192 LOCUS	AX093192 396 bp DNA PAT 30-MAR-2001
DEFINITION	Sequence 10 from Patent WO0118046.
ACCESSION	AX093192 GI:13509641
VERSION	AX093192.1 GI:13509641
KEYWORDS	
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 396) Xu,J. and Stolk,J.A.

TITLE	Ovarian tumor sequences and methods of use therefor
JOURNAL	Patent: WO 0118046-A 10-15-MAR-2001;
CORIXA CORPORATION (US)	
FEATURES	Location/Qualifiers
SOURCE	1..396 /organism="Homo sapiens" /db_xref="taxon:9606"
Misc-feature	1..396 /note="n = A,T,C or G"
BASE COUNT	85 a 18 c 55 g 153 t 85 others
ORIGIN	
Query Match	12.0%; Score 76; DB 6; Length 396; Best Local Similarity 78.6%; Pred. No. 0.00026; Matches 88; Conservative 0; Mismatches 24; Indels 0; Gaps 0;
Dy	7 tttcttttttttttttttttttttttttttttttttggtgggaaacctttt 66 Db 22 TTTT TT 81
Dz	67 tttctttttttttttttttaaacaccccccattatcattagaga 118 Db 82 TTTT TTTTTTTTTTTTTTTTTTTAAAAAAANNNNTTTTTTTTTTNAAA 133
RESULT 12	
LOCUS	AKO26528 1483 bp mRNA PRI 29-SEP-2000
DEFINITION	Homo sapiens CDNA: FLJ22875 fs, clone KAT02879.
ACCESSION	AKO26528
VERSION	AKO26528.1 GI:10439405
KEYWORDS	oligo cloning; fis (full insert sequence).
SOURCE	Homo sapiens signet-ring cell carcinoma cell_line:KATO III CDNA to mRNA, clone_1lb:KAT clone:KAT02879. Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo. 1 (sites) Matanabe,K., Kumagai,A., Itakura,S., Yamazaki,M., Tashiro,H., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T., Shibahara,T., Tanaka,T., Nakamura.Y., Isogai.T. and Sugano.S. NEDO human cDNA sequencing project Unpublished (2000) 2 (bases 1 to 1483) Sugano,S., Suzuki,Y., Ota,T., Obayashi.M., Nishi.T., Isogai.T., Shibahara.T., Tanaka,T. and Nakamura,Y. Direct Submission Submitted (29-AUG-2000) to the DDBJ/EMBL/GenBank databases. Sumio Sugano, Institute of Medical Science, University of Tokyo, Laboratory of Genome Structure Analysis, Human Genome Center; Shirokane-dai, 4-6-1, Minato-Ku, Tokyo 108-8639, Japan (E-mail:cna@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286, Fax:81-3-5449-5416)
COMMENT	NEDO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan; CDNA full Insert sequencing: Research Association for Biotechnology; CDNA library construction, 5'- & 3'-end one pass sequencing: Department of Virology and Human Genome Center, Institute of Medical Science, University of Tokyo (partly supported by Science and Technology Agency). Location/Qualifiers 1..1483 /organism="Homo sapiens" /db_xref="taxon:9606" /cell_line="KATO III" /cell_type="signet-ring cell carcinoma" /clone="KAT02879" /clone_id="KAT" /note="Cloning vector pWEISFLJ3" 31..513 /note="unnamed protein product" /codon_start=1
FEATURES	
SOURCE	
CDS	

					/protein_id="BAB15496.1"
					/db_xref="GI:10439406"
					/translation="MQRVSGLSMTPLSNVLMGLSEFGAARQPRIMEKALAEYDLI
					RITIDPFRPTLLEEVAVSCEVOEINIEBEYLVIIRFTPTPHCSLATLIGLCRLR
					KIORCLPFKHKLKEIYISEGHTSTEDINKQINDKERVAAMENPNLRIVEQCVLEPE
					"
BASE COUNT	519 a	272 c	308 g	384 t	
ORIGIN					
Query Match	12.0%;	Score 76;	DB 9;	Length 1483;	
Best Local Similarity	78.4%;	Pred. No. 0.00027;			
Matches	91;	Conservative	0;	Mismatches .25;	Indels 0;
Gaps					0.
Oy	3	aagcttggggaacct	62		
Dy	965	AAGGTATTTTATTT	906		
Oy	63	ttttttttttttttttttttttttaaaaccccccttatatttaagacaga	118		
Dy	905	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTAAAGATCATCTGCATTATTTATTTACAGA	850		
RESULT 13					
AX187054					
LOCUS	AX187054	612 bp	DNA		PAT 06-AUG-2001
DEFINITION	Sequence 2749 from Patent WO0142467.				
ACCESSION	AX187054				
VERSION	AX187054.1	GI:15138498			
KEYWORDS					
SOURCE	human.				
ORGANISM	Homo sapiens				
	Eumetazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
	Mammalia; Metazoa; Primates; Catarrhini; Hominidae; Homo.				
	1. (bases 1 to 612)				
REFERENCE	Schlegel,R., Deeds,J., Berger,A. and Zhao,X.				
AUTHORS	Genes, compositions, kits, and methods for identification,				
TITLE	assessment, prevention, and therapy of cervical cancer				
JOURNAL	Patent: WO 0142467-A 2749 14-JUN-2001;				
	Millennium Predictive Medicine, Inc. (us)				
FEATURES	Location/Qualifiers				
source	1..612				
	/organism="Homo sapiens"				
	/db_xref="taxon:9606"				
BASE COUNT	225 a	61 c	93 g	232 t	1 others
ORIGIN					
Query Match	12.0%;	Score 75.6;	DB 6;	Length 612;	
Best Local Similarity	69.9%;	Pred. No. 0.00031;			
Matches	102;	Conservative	0;	Mismatches 44;	Indels 0;
Gaps					0.
Oy	7	tttgaggaaaacttttc	66		
Dy	6	TTTTTTTTTTTTTTTTTTTTTTTTTCCTTTTTTTTTTTTTTTTTTTTTTTTT	65		
Oy	67	ttttttttttttttttttttttaaacacccttatatttaagacagcagccccc	126		
Dy	66	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTAAAAAATAAAAAAA	125		
Oy	127	acaaaaaaggaggcggtcggaana	152		
Dy	126	ACCAAAATAAAGGGGGCGGAATAAAA	151		
RESULT 14					
AK000614/c					
LOCUS	AK000614	1163 bp	mRNA		PRI 22-FEB-2000
DEFINITION	Homo sapiens CDNA FLJ20607 fls, clone KAT06217.				
ACCESSION	AK000614				
VERSION	AK000614.1	GI:7020825			
KEYWORDS	Oligo clapping; fls (full insert sequence).				
SOURCE	Homo sapiens signet-ring cell carcinoma cell_line:KATO III CDNA to				

[illegible]

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Comugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 17, 2002, 05:58:24 ; Search time 93.06 Seconds
(without alignments)
255,536 Million cell updates/sec

Title: US-09-823-101-7

Perfect score: 105
Sequence: 1 tgcagtggtccagacatcc.....ggccgctctctgaagctt 105

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 11328999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A.COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B.COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PTUS.COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Query Length	DB ID	Description
1	52.6	50.1	383	4 US-09-079-372-13	Sequence 13, Appl
2	46.6	44.4	613	4 US-09-079-372-1	Sequence 1, Appl
3	25.2	24.0	3483	3 US-08-808-599A-23	Sequence 23, Appl
4	25	23.8	1293	3 US-09-182-816-27	Sequence 27, Appl
5	25	23.8	1293	3 US-09-471-528-27	Sequence 27, Appl
6	25	23.8	1293	4 US-09-634-530-27	Sequence 27, Appl
7	25	23.8	1395	3 US-09-182-816-25	Sequence 25, Appl
8	25	23.8	1395	3 US-09-182-816-26	Sequence 26, Appl
9	25	23.8	1395	3 US-09-471-528-25	Sequence 25, Appl
10	25	23.8	1395	3 US-09-471-528-26	Sequence 26, Appl
11	25	23.8	1395	4 US-09-634-530-25	Sequence 25, Appl
12	25	23.8	1395	4 US-09-634-530-26	Sequence 26, Appl
13	25	23.8	1736	3 US-09-182-816-22	Sequence 22, Appl
14	25	23.8	1736	3 US-09-182-816-24	Sequence 24, Appl
15	25	23.8	1736	3 US-09-471-528-22	Sequence 22, Appl
16	25	23.8	1736	3 US-09-471-528-24	Sequence 24, Appl
17	25	23.8	1736	4 US-09-634-530-22	Sequence 22, Appl
18	25	23.8	1736	4 US-09-634-530-24	Sequence 24, Appl
19	24.8	23.6	4092	4 US-08-469-537A-106	Sequence 106, App
20	24.8	23.6	4547	4 US-09-723-534-3	Sequence 3, Appl
21	24.2	23.0	905	2 US-08-949-603-1	Sequence 1, Appl
22	24.2	23.0	905	2 US-08-706-370A-1	Sequence 1, Appl
23	24.2	23.0	905	2 US-08-949-580-1	Sequence 1, Appl
24	24.2	23.0	905	2 US-08-950-172A-1	Sequence 1, Appl
25	24.2	23.0	2268	3 US-09-344-579-1	Sequence 1, Appl
26	24.2	23.0	152331	3 US-09-128-155-16	Sequence 16, Appl
27	24.2	23.0	246240	2 US-08-724-394A-20	Sequence 20, Appl

28	24.2	23.0	246240	2 US-08-724-394A-21	Sequence 21, Appl
29	24.2	23.0	246240	2 US-08-724-394A-22	Sequence 22, Appl
30	24	22.9	14060	3 US-08-658-136-4	Sequence 4, Appl
31	24	22.9	53526	3 US-08-658-136-2	Sequence 2, Appl
32	24	22.9	53577	3 US-08-658-136-1	Sequence 1, Appl
33	23.6	22.5	1743	2 US-08-591-079-1	Sequence 1, Appl
34	23.6	22.5	2297	1 US-08-056-200-110	Sequence 110, App
35	23.6	22.5	2297	2 US-08-800-644-110	Sequence 110, App
36	23.6	22.5	35100	1 US-08-306-691B-19	Sequence 19, App
37	23.6	22.5	35100	5 PCT-US93-06251-19	Sequence 19, Appl
38	23.6	22.5	56516	2 US-08-996-306-1	Sequence 1, Appl
39	23.6	22.5	56516	4 US-09-338-907-1	Sequence 1, Appl
40	23.6	22.5	56520	4 US-09-338-907-179	Sequence 179, App
41	23.4	22.3	6873	4 US-09-131-028A-1	Sequence 1, Appl
42	23.4	22.3	6873	4 US-09-131-028A-8	Sequence 8, Appl
43	23.4	22.3	8430	4 US-09-131-028A-6	Sequence 6, Appl
44	23.4	22.3	8430	4 US-09-131-028A-10	Sequence 10, Appl
45	23.2	22.1	1960	2 US-08-533-306A-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-09-079-372-13

Sequence 13, Application US/09079372
Patent No. 616590

GENERAL INFORMATION:
APPLICANT: Singh, Pomila

APPLICANT: Wood, T.

TITLE OF INVENTION: INHIBITION OF ENDOGENOUS GASTRIN
EXPRESSION FOR TREATMENT OF COLORECTAL CANCER

NUMBER OF SEQUENCES: 17

CORRESPONDENCE ADDRESS:
ADDRESS: Arnold, White & Durkee

STREET: P.O. Box 4433
CITY: Houston

STATE: Texas
COUNTRY: United States of America

ZIP: 77210

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/079,372

FILING DATE: Concurrently Herewith
CLASSIFICATION:
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/634,546
FILING DATE: 18-APR-1996

ATTORNEY/AGENT INFORMATION:
NAME: Hodgins, Daniel S.

REGISTRATION NUMBER: 31,026
INFORMATION FOR SEQ ID NO: 13:

SEQUENCE CHARACTERISTICS:
LENGTH: 383 base pairs

TYPE: nucleic acid
STRANDEDNESS: single

TOPOLOGY: linear

US-09-079-372-13

Query Match 50.1% Score 52.6; DB 4; Length 383;
Best Local Similarity 82.8%; Pred. No. 2.9e-10;

Matches 72; Conservative 0; Mismatches 14; Indels 1; Gaps 1;

QY 19 cctgcagcttctgttaccatgatcagcagcagatgtgtatgtctctctgcact 78
||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 24 cctccacactctctgcagcagcagatgcacacgactatgtctgtatgtcgtgactttgcact 83

```

OY 79 agatcagccgcctctctgaagctt 105
    ||||| ||||| ||||| ||||| |||||
Db 84 -GGCTCTGGCCCGCCTTCTCTGAAGCTT 109

RESULT 2
US-09-079-372-1
: Sequence 1, Application US/09079372
: Patent No. 6165990
: GENERAL INFORMATION:
: APPLICANT: Singh, Pomila
: APPLICANT: Wood, T.
: TITLE OF INVENTION: INHIBITION OF ENDOGENOUS GASTRIN
: TITLE OF INVENTION: EXPRESSION FOR TREATMENT OF COLORECTAL CANCER
: NUMBER OF SEQUENCES: 17
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Arnold, White & Durkee
: STREET: P.O. Box 4433
: CITY: Houston
: STATE: Texas
: COUNTRY: United States of America
: ZIP: 77210
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/079,372
: FILING DATE: Concurrently Herewith
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/634,546
: FILING DATE: 18-APR-1996
: ATTORNEY/AGENT INFORMATION:
: NAME: Hodgins, Daniel S.
: REGISTRATION NUMBER: 31,026
: REFERENCE/DOCKET NUMBER: UTSG:220
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 613 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
:
US-09-079-372-1

Query Match 44.4%; Score 46.6; DB 4; Length 613;
Best Local Similarity 92.3%; Fred. No. 5.2e-08;
Matches 60; Conservative 0; Mismatches 4; Indels 1; Gaps 1.

OY 41 atcagcagcagctgtgtatgtctgtccttgacacagatcagagccgcctctctga 100
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1 ATGCAGCAGCATATGTGTATGTATGTCTGATCTTGCACT-GGCTCTGGCCCGCCTTCTCTGA 59
    |||||
OY 101 agctt 105
    |||||
Db 60 AGCTT 64

RESULT 3
US-08-808-599A-23
: Sequence 23, Application US/08808599A
: Patent No. 6111089
: GENERAL INFORMATION:
: APPLICANT: Fukuda, Michiko N.
: TITLE OF INVENTION: Trophinin, Trophinin-Assisting
: TITLE OF INVENTION: Protein and Methods to Inhibit Implantation
: NUMBER OF SEQUENCES: 41
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Campbell & Flores LLP
: STREET: 4370 La Jolla Village Drive, Suite 700
: CITY: San Diego

```

```

STATE: California
COUNTRY: USA
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/808,599A
FILING DATE: 28-FEB-1997
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/317,522
FILING DATE: 04-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/439,818
FILING DATE: 12-MAY-1995
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-LA 2256
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 23:
SEQUENCE CHARACTERISTICS:
LENGTH: 3483 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 1..3481
US-08-808-599A-23

Query Match 24.0%; Score 25.2; DB 3; Length 3483;
Best Local Similarity 57.7%; Pred. No. 5.7;
Matches 45; Conservative 0; Mismatches 33; Indels 0; Gaps

QY 9 ttcagagatctctgcacggtctgtgatgatgcagcagcagatgtgtatgtctgc 68
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 672 TGGCAGCACCTAGTACTAGTGCCTTCTGTACGCAGCAGCATTAGCTTGTGTGTC 731
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

OY 69 tcttcagcactagatcag 86
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 732 ACCCAGCAGCTAGCAGCAG 749

RESULT 4
US-09-182-816-27/C
Sequence 27, Application US/09182816
Patent No. 6143542
GENERAL INFORMATION:
APPLICANT: Wisniewski, Nancy
APPLICANT: Silver, Gary M.
APPLICANT: Lo, Katherine C.
APPLICANT: Brandt, Kevin S.
TITLE OF INVENTION: NOVEL FLEA EPOXIDE HYDROLASE NUCLEIC ACID MOLECULES
FILE REFERENCE: FC-3-C1
CURRENT APPLICATION NUMBER: US/09/182,816
CURRENT FILING DATE: 1998-10-29
EARLIER APPLICATION NUMBER: 08/989,510
EARLIER FILING DATE: 1997-12-12
NUMBER OF SEQ ID NOS: 31
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 27
LENGTH: 1293
TYPE: DNA
ORGANISM: Ctenocephalides felis
FEATURE:

```



```
; GENERAL INFORMATION:
; APPLICANT: Wisniewski, Nancy
; APPLICANT: Silver, Gary M.
; APPLICANT: Lo, Katherine C.
; APPLICANT: Brandt, Kevin S.
; TITLE OF INVENTION: NOVEL FLEA EPOXIDE HYDROLASE NUCLEIC ACID MOLECULES,
; FILE REFERENCE: FC-3-C1
; CURRENT APPLICATION NUMBER: US/09/182,816
; EARLIER FILING DATE: 1998-10-29
; EARLIER APPLICATION NUMBER: 08/989,510
; EARLIER FILING DATE: 1997-12-12
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 26
; LENGTH: 1395
; TYPE: DNA
; ORGANISM: Ctenocephalides felis
US-09-182-816-26
```

```
Query Match          23.8%; Score 25; DB 3; Length 1395;
Best Local Similarity 58.9%; Pred. No. 5;
Matches 43; Conservative 0; Mismatches 30; Indels 0; Gaps 0;
```

```
OY 29 ctgtgtacgatgacgcagcagatgtgtgtatgtgtcttcttgactagatcagc 88
      ||||| | | | | | | | | | | | | | | | | | | | | | | | |
DB 480 ctgtgcaggtttgttagctgtgatgataatcctctgtcttccaaaagggtgaaga 539
OY 89 cgccctctctgaa 101
      | | | | | | | |
DB 540 agtcttcgtcctcaa 552
```

```
RESULT 9
US-09-471-528-25/C
; Sequence 25, Application US/09471528
; Patent No. 6153397
; GENERAL INFORMATION:
; APPLICANT: Wisniewski, Nancy
; APPLICANT: Silver, Gary M.
; APPLICANT: Lo, Katherine C.
; APPLICANT: Brandt, Kevin S.
; TITLE OF INVENTION: FLEA EPOXIDE HYDROLASE PROTEINS AND USES THEREOF
; FILE REFERENCE: FC-3-C1-1
; CURRENT APPLICATION NUMBER: US/09/471,528
; CURRENT FILING DATE: 1999-12-27
; EARLIER APPLICATION NUMBER: 09/182,816
; EARLIER FILING DATE: 1998-10-29
; EARLIER APPLICATION NUMBER: 08/989,510
; EARLIER FILING DATE: 1997-12-12
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 25
; LENGTH: 1395
; TYPE: DNA
; ORGANISM: Ctenocephalides felis
US-09-471-528-25
```

```
Query Match          23.8%; Score 25; DB 3; Length 1395;
Best Local Similarity 58.9%; Pred. No. 5;
Matches 43; Conservative 0; Mismatches 30; Indels 0; Gaps 0;
```

```
OY 29 ctgtgtacgatgacgcagcagatgtgtgtatgtgtcttcttgactagatcagc 88
      ||||| | | | | | | | | | | | | | | | | | | | | | | | |
DB 916 ctgtgtcaggtttgttagctgtgatgataatcctctgtcttccaaaagggtgaaga 857
OY 89 cgccctctctgaa 101
      | | | | | | | |
DB 856 agtcttcgtcctcaa 844
```

```
RESULT 10
US-09-471-528-26
; Sequence 26, Application US/09471528
; Patent No. 6153397
; GENERAL INFORMATION:
; APPLICANT: Wisniewski, Nancy
; APPLICANT: Silver, Gary M.
; APPLICANT: Lo, Katherine C.
; APPLICANT: Brandt, Kevin S.
; TITLE OF INVENTION: FLEA EPOXIDE HYDROLASE PROTEINS AND USES THEREOF
; FILE REFERENCE: FC-3-C1-1
; CURRENT APPLICATION NUMBER: US/09/471,528
; CURRENT FILING DATE: 1999-12-27
; EARLIER APPLICATION NUMBER: 09/182,816
; EARLIER FILING DATE: 1998-10-29
; EARLIER APPLICATION NUMBER: 08/989,510
; EARLIER FILING DATE: 1997-12-12
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 26
; LENGTH: 1395
; TYPE: DNA
; ORGANISM: Ctenocephalides felis
US-09-471-528-26
```

```
Query Match          23.8%; Score 25; DB 3; Length 1395;
Best Local Similarity 58.9%; Pred. No. 5;
Matches 43; Conservative 0; Mismatches 30; Indels 0; Gaps 0;
```

```
OY 29 ctgtgtacgatgacgcagcagatgtgtgtatgtgtcttcttgactagatcagc 88
      ||||| | | | | | | | | | | | | | | | | | | | | | | | |
DB 480 ctgtgcaggtttgttagctgtgatgataatcctctgtcttccaaaagggtgaaga 539
OY 89 cgccctctctgaa 101
      | | | | | | | |
DB 540 agtcttcgtcctcaa 552
```

```
RESULT 11
US-09-634-530-25/C
; Sequence 25, Application US/09634530
; Patent No. 6290958
; GENERAL INFORMATION:
; APPLICANT: Wisniewski, Nancy
; APPLICANT: Silver, Gary M.
; APPLICANT: Lo, Katherine C.
; APPLICANT: Brandt, Kevin S.
; TITLE OF INVENTION: FLEA EPOXIDE HYDROLASE PROTEINS AND USES THEREOF
; FILE REFERENCE: FC-3-C1-1
; CURRENT APPLICATION NUMBER: US/09/634,530
; CURRENT FILING DATE: 2000-08-08
; PRIOR APPLICATION NUMBER: 09/471,528
; PRIOR FILING DATE: 1999-12-27
; PRIOR APPLICATION NUMBER: 09/182,816
; PRIOR FILING DATE: 1998-10-29
; PRIOR APPLICATION NUMBER: 08/989,510
; PRIOR FILING DATE: 1997-12-12
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 25
; LENGTH: 1395
; TYPE: DNA
; ORGANISM: Ctenocephalides felis
US-09-634-530-25
```

```
Query Match          23.8%; Score 25; DB 4; Length 1395;
```

```
Best Local Similarity 58.9%; Pred. No. 5;
Matches 43; Conservative 0; Mismatches 30; Indels 0; Gaps 0;
OY 29 ctgtgtacgatgacgcagcagatgtgtgtatgtgtcttcttgactagatcagc 88
      ||||| | | | | | | | | | | | | | | | | | | | | | | | |
```


; SEQ ID NO 22
; LENGTH: 1736
; TYPE: DNA
; ORGANISM: Ctenocephalides felis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (159)..(1553)
US-09-471-528-22

Query Match 23.8%; Score 25; DB 3; Length 1736;
Best Local Similarity 58.9%; Pred. No. 5.3;
Matches 43; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

Qy 29 ctgtgacgatagcagcagcagatgtgtatgtctgtctcttgactgagcagc 88
Db 1074 CTGTGTCAGCTTTGTAGCTTGATGGAATATCTCTGATCTCCAAAGGCTGAGA 1015

Qy 89 cgccttctctgaa 101
Db 1014 AGTGTGCTCAA 1002

Search completed: January 17, 2002, 05:58:26
Job time: 1034 sec

CC tissues and cells the genes are expressed in. Example of activities
 CC include: cytostatic; proliferative; vunerary; immunomodulator;
 CC antidiabetic; antisthmatic; antirheumatic; antirheumatic;
 CC antiinflammatory; antihypertoid; antiallergic; antibacterial; antiviral;
 CC dermatological; neuroprotective; cardiac; thrombolytic; coagulant;
 CC nocotropic; vasotropic; antipsoriatic and antiangiogenic. The
 CC polynucleotides and polypeptides can be used for preventing, treating or
 CC ameliorating medical conditions and diagnosing pathological conditions.
 CC Polynucleotides, polypeptides, antibodies, agonists and antagonists from
 CC the present invention may be used to treat immune disorders by activating
 CC or inhibiting the proliferation, differentiation or mobilisation of
 CC immune cells, to treat disorders of haematopoietic cells, autoimmune
 CC disorders, allergic reactions, graft versus host disease and organ
 CC rejection, modulate haemostatic or thrombolytic activity, modulate
 CC inflammation, cancers, cardiovascular disorders, neurological disease and
 CC bacterial or viral infections. The peptides, nucleotides, antibodies,
 CC agonists and antagonists may be also be used in drug screens. AAC78449 to
 CC MAC78457 and AAB44240 represent sequences used in the exemplification of
 CC the present invention.

XX
 SQ Sequence 179 AA;

alignment_scores:
 Quality: 80.00 Length: 25
 Ratio: 3.810 Gaps: 0
 Percent Similarity: 84.000 Percent Identity: 64.000

alignment_block:

US-09-823-101-1 x AAB44027 ..

Align seg 1/1 to: AAB44027 from: 1 to: 179

464 AAGAGAAAGAACTACTCAAGGTTAAACCAAAAAAAGAAAAA 513
 ||||| : : : : : ||||| : : : : : ||||| : : : : :
 2 LysLysLys***LysLysLysLysLysLysLysLysLysLys 18
 514 AAAAAAGGGGGGGAACACAGGCGC 538
 ||||| : : : : : ||||| : : : : : ||||| : : : : :
 18 sLysLysGlyGlyArg****Gly 26

seq_name: /SIDS8/gcgdata/geneseq/geneseq/AA1999.DAT:AAW88308

seq_documentation_block:

ID AAW88308 standard; Protein: 350 AA.

AC AAW88308;

DT 26-APR-1999 (first entry)

DE E. coli O antigen polymerase.

XX O antigen: O111 antigen: wzv gene; polymerase: diarrhoea;

KM haemorrhagic colitis; diagnosis.

XX Escherichia coli.

OS MO9850531-A1.

PN 12-NOV-1998.

PD 01-MAY-1998; 98WO-AU00315.

PR 22-JUL-1997; 97AU-0008162.

PR 01-MAY-1997; 97AU-0006545.

XX (UNSY) UNIV STDNEY.

PA Reeves PR, Wang L;

XX WPI: 1999-059669/05.

DR N-PSDB: AAX06748.

PT Nucleic acid molecules specific for bacterial polysaccharide
 PT antigens - useful for detecting specific strains in, e.g. food,
 PT faeces or patient samples

PS Disclosure; Fig 7; 165pp; English.

CC This is the amino acid sequence of an O antigen polymerase encoded
 CC by the wzv gene of a gene cluster (see AAX06748) involved in the
 CC biosynthesis of the Escherichia coli O111 O antigen. The protein
 CC has 10 predicted transmembrane segments and a large cytoplasmic
 CC loop. This inner membrane topology is a characteristic feature
 CC of all known O antigen polymerases. The use of nucleic acid
 CC molecules derived from particular assembly and transport genes,
 CC particularly wbd (transferase), wzx (flippase) and wzy (polymerase)
 CC genes, within O antigen gene clusters improves the specificity of
 CC methods for the detection and identification of O antigens, e.g. in
 CC testing food- or faecal-derived samples, or samples from patients.
 CC The O antigen is a major virulence factor of enteropathogenic E.
 CC coli strains that cause diarrhoea and haemorrhagic colitis.

XX
 SQ Sequence 350 AA;

alignment_scores:
 Quality: 80.00 Length: 229
 Ratio: 0.816 Gaps: 10
 Percent Similarity: 42.795 Percent Identity: 22.707

alignment_block:

US-09-823-101-1/rev x AAW88308 ..

Align seg 1/1 to: AAW88308 from: 1 to: 350

506 TTTTGTGTTTGTGTTTAACTTGAAGTGTGTTCTTC.....TT 463
 ||| ||||| : : : : : ||||| : : : : : ||||| : : : : :
 98 PheAsnPheLeuTYrLeuLysThrPheAspLysPheSerSerPheProph 114
 462 TAGTATCTTGTTCATTATGACATATTTCTGACCTGGACAGGGAATT 414
 || : : : : : || : : : : : || : : : : : || : : : : :
 114 eThrPhePheIleLeuPheTYrLeuPheArgLeuGlyIleGlyAsnL 131
 413 A.....AAAACCAATTTTGTAGTCAGAAATCTTAAAGTAT 373
 ||||| : : : : : ||||| : : : : : ||||| : : : : :
 131 euProValTYrLysAsn.....LysLys 138
 372 TTTCTGTCCTTATATAAATTTGAAATATAAAGAACAAAGATTCAC 323
 ||| ||||| : : : : : ||||| : : : : : ||||| : : : : :
 139 PheTYrAlaLeuIlePheLeuPheIleLeuIleAspIleMetGlnSerLe 155
 322 CATATTCACGACGACGACAAATGCTGCTGACTATTTGGCA..... 280
 :
 155 uLeuIleAsnTYrArgGlyGlnIleLeuTYrSerValIleGlyIleLeuI 172
 279TATTTC.....TATTTC..... 274
 172 IeLeuValPheLysValAsnLeuArgLysLysIleProTYrPhePheLeu 188
 273 ...CTTCGTCCTTTTCTGCTGCAATATGAA...ACTGAGAGCTGGA 230
 ||||| : : : : : ||||| : : : : : ||||| : : : : :
 189 MetLeuProValLeuTYrValIleIleMetAlaTYrIleGlyPheAsnTY 205
 229 TTTTTCAGTGCACGACGACATGAGTGGACCTACACAGAA..... 190
 || : : : : : || : : : : : || : : : : : || : : : : :
 205 rPheAsnLysGlyValThrPhePheGluProThrAlaSerAsnIleGlyA 222
 189TGTTCTGTTTC 178
 222 rGThGlyMetIleTYrTYrLeuValSerGlnLeuGlyAspTYrIlePhe 238
 177 CAT.....ATTAGGT 167
 ||| : : : : : ||| : : : : : ||| : : : : :
 239 HisGlyMetGlyThrLeuAsnPheLeuAsnAsnGlyGlyGlnTYrLysTh 255

DR WPI: 2001-442253/47.
 DR N-PSDB: AAI60659.
 PT Novel nucleic acids and polypeptides, useful for treating disorders
 PT such as central nervous system injuries -
 XX
 XX Example 2: SEQ ID NO 6434; 10078bp; English.
 XX
 CC The invention relates to human nucleic acids (AA157798-AA161369) and
 CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: Immune system suppression,
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukaemias and
 CC C.N.S disorders.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification.
 CC
 XX
 XX Sequence 654 AA;
 SO

alignment_scores:
 Quality: 78.00 Length: 161
 Ratio: 0.987 Gaps: 7
 Percent Similarity: 49.068 Percent Identity: 24.224

alignment_block:
 US-09-823-101-1 x AAM41503 ..
 Align seg 1/1 to: AAM41503 from: 1 to: 654

100 GAAAAAGAGAAAAAGAGAGTATGAGGAAAGTACGACGACAGAG 149
 ||||| |||||:||||:||||: |||||
 141 GtutysgllysgluluglulysmetglulysThraspoluh 157
 :||| |||||:||||: |||||
 150 CAGG...CAGAGACCAATGACCTTAATATGAAAAACAGACATTCCTGT 196
 :||| |||||:||||: |||||
 157 rargleuglulysglulaglulpheluglulysSeratlatysgluasleua 174
 :||| |||||:||||: |||||
 197 GAGAGCCATCATGCTCAGTGCACACTGAAAAATCCAGCTCCCACTTT 246
 ||| :|||:||||: |||||
 174 rGAsperlysgluleuarGAsnPhelglalaleuglulileaspaplle 190
 :||| :|||:||||: |||||
 247 CTATATGACAGAAAAAGAGAGAAATATGCCAAAT..... 288
 :||| :|||:||||: |||||
 191 MetAlaIlelysmetCluAspProllysglulilearGlysglulucel 207
 :||| :|||:||||: |||||
 289AGTCACAGTGAATTTGTCTGAGTGTGAATTATG 322
 :||| :|||:||||: |||||
 207 uCluAspHislyscysSerHis..... 214
 :||| :|||:||||: |||||
 322 GCTGATTCCTTTGTTCTTTATATTTCAAAATTTTATAAGACACAGAA 372
 :||| :|||:||||: |||||
 215PheProAspHeserlySeratlaser 223
 :||| :|||:||||: |||||
 373 AATCTTTTAAAGAAATTTGACTACAAATAATGCTTTTAAATTCCTGT 422
 :||| :|||:||||: |||||
 224 SerlyIlelelleSerasp.....Va 231
 :||| :|||:||||: |||||
 423 GCCAAGT.....TCAGAAATATGCCAT...AATGACACAGATACTAA 463
 :||| :|||:||||: |||||
 231 lProSeratGlylsAspHismetCysHisProHisglYIlelelleleG 248
 :||| :|||:||||: |||||
 464 AAGAAAGAAATACCTC.....AAGGTAATAAAAAACCAAAAAA 501
 :||| :|||:||||: |||||
 248 lAspProAlaIleuAsnlysProGluIuIeulyslyslYslYslYsls 264

seq_name: /SID58/gcgcdata/geneseq/geneseqp/AA2001.DAT: AAB94440
 502 AAAAAAAAAAAAAAAAAAGGGGGGGAACACAG 534
 |||||:||||: |||||:||||: |||||
 265 LysSerlyMetAspArgHlsGlyAsnAspLys 275

seq_documentation_block:
 ID AAB94440 standard; Protein: 245 AA.
 AC AAB94440;
 DT 26-JUN-2001 (first entry)
 DE Human protein sequence SEQ ID NO:15068.
 DE
 KW Human; primer: detection; diagnosis; antisense therapy; gene therapy.
 OS Homo sapiens.
 XX EPI074617-A2.
 XX
 PD 07-FEB-2001.
 XX
 PF 28-JUL-2000; 2000EP-0116126.
 XX
 PR 29-JUL-1999; 99JP-0248036.
 PR 27-AUG-1999; 99JP-0300253.
 PR 11-JAN-2000; 2000JP-0118776.
 PR 02-MAY-2000; 2000JP-0183767.
 PR 09-JUN-2000; 2000JP-0241899.
 XX
 PA (HELI-) HELIX RES INST.
 XX
 PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 DR WPI: 2001-318749/34.
 XX
 XX
 PT Primer sets for synthesizing polynucleotides, particularly the 5602
 PT full-length cDNAs defined in the specification, and for the detection
 PT and/or diagnosis of the abnormality of the proteins encoded by the
 PT full-length cDNAs -
 XX
 PS Claim 8: SEQ ID 15068; 2537bp + CD ROM; English.
 XX
 CC The present invention describes primer sets for synthesizing 5602
 CC full-length cDNAs defined in the specification. Where a primer set
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
 CC AAB95883 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.
 XX
 SO Sequence 245 AA;
 alignment_scores:

Quality: 77.50 Length: 91
 Ratio: 1.409 Gaps: 4
 Percent Similarity: 60.440 Percent Identity: 26.374

alignment_block:

US-09-823-101-1 x AAB94440 ..

Align seg 1/1 to: AAB94440 from: 1 to: 245

```

14 ACCGAGATCGAGTGAAGACTGACATTATGCTAACCAAGACGACACTCC 63
   ||| |||.....:||||| |||:
3  ThrValAspIleAlaAsnSerThrLeuGlyLys.....Pr 15
64 TAAAGTTAAGTGAAGCGAGGCTAGAACAAATAAGCAAAAAGAAAAA 113
   ||| |||.....:||||| |||:
15 Olys.....ArgLysArgArgLysL 22
114 AGACAGGTCATTAGTGGGGAAGGTAGCCAGACGACGACGACAC 163
   ||| |||.....:||||| |||:
22 yslYsGlyHISgLyTrpSerArgThrGlyThrArgGlnLysAsn 38
164 ATGACCTTAATGCAAAACAGAACTTCTGTAGTCCACTCATGCT 213
   ||| |||.....:||||| |||:
39 Gln.....GlnAsnAspAsnSerLysAlaAspGlyGlnLeuVa 51
214 CAGTGCCACTGAAATATCCAGCTCTCCAGTTTCTATATGCGAGAAA 263
   ||| |||.....:||||| |||:
51 lSerSerGlyLysLysAlaAsnMetAsnLeuLysAspLeuSerIleA 68
264 GGACAGGAGAGGAAATATGCCAA 286
   || ||| |||.....:|||||
68 rg...GlyArgLysArgGlyLys 74

```

seq_name: /SIDs8/gcgdata/geneseq/geneseqp/AA2001.DAT:AA674650

seq_documentation_block:

ID AAG74650 standard; Protein: 69 AA.

AC AAG74650;

DT 03-SEP-2001 (first entry)

DE Human colon cancer antigen protein SEQ ID NO:5414.

KW Human: colon cancer; colon cancer antigen; diagnosis; detection;

KX colorectal carcinoma.

OS Homo sapiens.

PN WO200122920-A2.

PD 05-APR-2001.

PF 28-SEP-2000: 2000WO-US26524.

PR 29-SEP-1999: 99US-0157137.

PR 03-NOV-1999: 99US-0163280.

XX (HUMA-) HUMAN GENOME SCI INC.

PA Ruben SM, Barash SC, Birse CE, Rosen CA;

DR WPI: 2001-235357/24.

DR N-PSDB: AAH34055.

XX Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
 PT useful for preventing, diagnosing and/or treating colorectal cancers -
 XX
 XX
 PS Claim 11: Page 7038; 9803pp; English.

CC AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
 CC cancer-associated nucleic acid molecules (N) and proteins (P), where
 CC the proteins are collectively known as colon cancer antigens. The colon

CC cancer antigens have cytostatic activity and can be used in gene
 CC therapy and vaccine production. N and P may be used in the prevention,
 CC diagnosis and treatment of diseases associated with inappropriate P
 CC expression. For example, N and P may be used to treat disorders
 CC associated with decreased expression by rectifying mutations or deletions
 CC in a patient's genome that affect the activity of P by expressing
 CC inactive proteins or to supplement the patient's own production of P.
 CC Additionally, N may be used to produce the colon cancer-associated Ps,
 CC by inserting the nucleic acids into a host cell and culturing the cell
 CC to express the proteins. N and P can be used in the prevention, diagnosis
 CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
 CC and AAB77789 represent sequences used in the exemplification of the
 CC present invention.
 CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
 CC missing at time of publication, meaning no sequences are present for
 CC SEQ ID NO:1027 to 1052, 7921 and 7922.

XX Sequence 69 AA;

alignment_scores:

Quality: 77.00 Length: 56
 Ratio: 2.200 Gaps: 2
 Percent Similarity: 62.500 Percent Identity: 37.500

alignment_block:

US-09-823-101-1 x AAG74650 ..

Align seg 1/1 to: AAG74650 from: 1 to: 69

```

371 AATATCTTTTAAGATTTGCTACTACAAATAATGTTTAAATTCCT 420
   ||| |||.....:||||| |||:
12 LysTyrPheIleGlnTyrValAlaPheValAsn...LeuThrAsnAspII 27
421 G.....TGCCAAAGTTTCAGAAATATGCAATATGCAAGACAG 455
   ||| |||.....:||||| |||:
27 eLysHISLysLeuGlnCysArgGlnLysLysLysLysLysLysLys 44
456 ATAACTAAAGAGAAACTACCTCAAGGTTAAATAACCAAAAAA 505
   ||| |||.....:||||| |||:
44 yslYsLysLysLysLysLysLysLysLysLysLysLysLysLys 60
506 AAAAAAAAAAAAAAGGCG 523
   ||| |||.....:||||| |||:
61 LysLysLysLysLysGly 66

```

seq_name: /SIDs8/gcgdata/geneseq/geneseqp/AA2000.DAT:AAB54269

seq_documentation_block:

ID AAB54269 standard; Protein: 122 AA.

AC AAB54269;

DT 09-MAR-2001 (first entry)

DE Human pancreatic cancer antigen protein sequence SEQ ID NO:721.

KW Human: pancreas; pancreatic cancer; pancreatic cancer antigen;
 KW detection; diagnosis; identification; cytostatic; neuroprotective;
 KW neoplastic; immunomodulatory; relaxant; contraceptive; gynaecological;
 KW antiinflammatory; cardiac; gene therapy; chromosome mapping;
 KW linkage analysis; tissue identification; tissue typing; forensic;
 KW neutral; immune system; muscular; reproductive; gastrointestinal;
 KW pulmonary; cardiovascular; renal; proliferative.

OS Homo sapiens.

PN WO200055320-A1.

PD 21-SEP-2000.

PF 08-MAR-2000: 2000WO-US05989.

XX


```
259 AAAAGCAGCAGGAGGAATATGCC.....AAATACTCA 293
    |||      ::| |
129 aValAlaThrLysCysGlnAlaIaThrThrGluGluAlaValAla 146
    :::::| |
294 CAGTGATTTGCTGAGTGGTGAATTATGGCTGATTTCTTCTTTTA 343
    ::||| |:::| |
146 laValThrLeuAlaLys.....AlaGluAlaMetAlaPheLeu 158
344 .....TATTTCCAAATTTT 357
    :::::| |
159 GlnGluGlnProPheLysAspPheValThrSerAlaPheTyrAspLysPh 175
358 TATAGA.....GCACAGAAATACTTTT 380
    :::::| |
175 eLeuGlnTTrPlyLeuPheGluMetGlnProValSerAspLysTyrPheT 192
381 AAGATTTCTGACTACAAAAATTGGTTTAAATCCCTGTGCCAAGTT 430
    ||||| ::| |
192 hrgIuPheArgValLeuGlyLysGlyGlyPheGlyGluValCysAlaVal 208
431 CAGAAATATGCCATAATGGAACAAGATACTAAAGAGAAACTACCTC 480
    ||| |:::| |
209 Gln.....ValLysAsnThrGln 214
481 AAGGTTA.....AAAAACCAAAAAAATAAAAAAAGG 521
    :::::| |
214 yLysMetTyrAlaCysLysLysLeuAspLysArgLeuLysLysG 231
522 GGGGGGAA 529
    |||||
231 yGlyGlyu 233
```

OM of: US-09-823-101-1 to: Issued_Patents_AA* out_format : pfs

Date: Jan 17, 2002 4:00 PM

About: Results were produced by the Gencore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODE=frame-n2p.model -DEV=xlp
-Q/cgn2_1/USPTO.spool/US09823101/runat.17012002.155137_22014/app-query.fasta_1.2239
-DB=Issued_Patents_AA -QPMF=fasten -SUFFIX=n2p.ra1
-GAPOP=12.000 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000
-LOOPEXT=0.000 -OGAPOP=4.500 -OGAPEXT=0.050 -XGAPOP=10.000
-XGAPEXT=0.500 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPO=10.000
-YGAPEXT=0.500 -DELOP=6.000 -DELEXT=7.000 -START=1
-MATRIX=bloms62 -TRANS=human40.cdi -LIST=45 -DOCLIGN=200
-THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pfs -NORM=ext -MINLEN=0 -MAXLEN=2000000000
-USER=US09823101 -ECGN1_1_59 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPHY
-WAIT -THREADS=1

Search information block:

Query: US-09-823-101-1
Query length: 542
Database: Issued_Patents_AA*
Database sequences: 212252
Database length: 22503292
Search time (sec): 66.610000

score_list:

Sequence	Strd Orig	ZScore	EScore	len	Documentation	
/cgn2_6/prodata/2/iaa/5b_COMB.pep:US-09-245-041-11		75.00	146.26	0.3063	351	1
/cgn2_6/prodata/2/iaa/5b_COMB.pep:US-09-245-041-19		73.50	133.97	0.5971	871	1
/cgn2_6/prodata/2/iaa/5b_COMB.pep:US-08-477-451-13		72.00	123.89	1.10	1724	1
/cgn2_6/prodata/2/iaa/5b_COMB.pep:US-08-221-817-13		70.50	131.21	1.29	576	1
/cgn2_6/prodata/2/iaa/5b_COMB.pep:US-08-454-439-13		70.50	131.21	1.29	576	1
/cgn2_6/prodata/2/iaa/5b_COMB.pep:US-08-942-686-2		69.50	126.32	1.84	753	1
/cgn2_6/prodata/2/iaa/5b_COMB.pep:US-08-737-078A-2		68.50	116.17	1.22	47	1
/cgn2_6/prodata/2/iaa/5b_COMB.pep:US-08-464-954A-6		68.50	126.67	2.31	576	1
/cgn2_6/prodata/2/iaa/5b_COMB.pep:US-08-477-451-3		68.50	109.98	3.56	3174	1
/cgn2_6/prodata/2/iaa/5b_COMB.pep:US-08-748-947A-2		67.00	125.76	3.34	446	1
/cgn2_6/prodata/2/iaa/5b_COMB.pep:US-09-068-650-5		66.00	147.07	2.43	40	1
/cgn2_6/prodata/2/iaa/5b_COMB.pep:US-09-612-126-11		66.00	132.42	3.55	179	1
/cgn2_6/prodata/2/iaa/5b_COMB.pep:US-09-612-126-8		66.00	132.04	3.58	186	1
/cgn2_6/prodata/2/iaa/5b_COMB.pep:US-09-612-126-1		66.00	129.58	3.88	255	1
/cgn2_6/prodata/2/iaa/5b_COMB.pep:US-08-997-803-15		65.50	129.58	4.29	213	1
/cgn2_6/prodata/2/iaa/5b_COMB.pep:US-08-748-506-12		65.50	125.57	4.76	321	1
/cgn2_6/prodata/2/iaa/5b_COMB.pep:US-08-748-506-20		65.50	125.57	4.76	321	1
/cgn2_6/prodata/2/iaa/5b_COMB.pep:US-08-836-786-4		64.50	147.50	3.40	27	1
/cgn2_6/prodata/2/iaa/5b_COMB.pep:US-08-152-488-8		64.50	145.84	3.55	32	1
/cgn2_6/prodata/2/iaa/5b_COMB.pep:US-08-303-025-8		64.50	145.84	3.55	32	1
/cgn2_6/prodata/2/iaa/5b_COMB.pep:US-08-677-304-8		64.50	145.84	3.55	32	1
/cgn2_6/prodata/2/iaa/5b_COMB.pep:US-09-183-959-12		64.50	125.21	6.06	264	1
/cgn2_6/prodata/2/iaa/5b_COMB.pep:US-08-477-451-6		64.50	100.92	11.39	3169	1
/cgn2_6/prodata/2/iaa/5b_COMB.pep:US-08-460-890A-58		64.00	141.38	4.47	45	1
/cgn2_6/prodata/2/iaa/5b_COMB.pep:US-08-167-641C-58		64.00	141.38	4.47	45	1
/cgn2_6/prodata/2/iaa/5b_COMB.pep:US-08-460-971A-58		64.00	141.38	4.47	45	1
/cgn2_6/prodata/2/iaa/5b_COMB.pep:US-08-462-040-58		64.00	138.73	4.79	59	1
/cgn2_6/prodata/2/iaa/5b_COMB.pep:US-08-462-040-60		64.00	138.73	4.79	59	1
/cgn2_6/prodata/2/iaa/5b_COMB.pep:US-08-167-641C-60		64.00	138.73	4.79	59	1
/cgn2_6/prodata/2/iaa/5b_COMB.pep:US-08-460-971A-60		64.00	138.73	4.79	59	1
/cgn2_6/prodata/2/iaa/5b_COMB.pep:US-08-462-040-60		64.00	103.07	12.10	2265	1
/cgn2_6/prodata/2/iaa/5b_COMB.pep:US-08-149-097D-36		64.00	102.06	12.41	2509	1
/cgn2_6/prodata/2/iaa/5b_COMB.pep:US-08-477-451-8		64.00	99.69	13.21	3200	1
/cgn2_6/prodata/2/iaa/5b_COMB.pep:US-08-221-817-22		63.50	115.32	9.88	576	1
/cgn2_6/prodata/2/iaa/5b_COMB.pep:US-08-454-439-22		63.50	115.32	9.88	576	1
/cgn2_6/prodata/2/iaa/5b_COMB.pep:US-08-959-104-11		63.50	113.88	10.26	667	1

/cgn2_6/prodata/2/iaa/5b_COMB.pep:US-08-477-451-7 - 63.50 98.29 15.38 3287
/cgn2_6/prodata/2/iaa/5b_COMB.pep:US-08-097-830E-2 + 63.00 143.07 5.40 30
/cgn2_6/prodata/2/iaa/5b_COMB.pep:US-08-456-112B-2 + 63.00 143.07 5.40 30
/cgn2_6/prodata/2/iaa/5b_COMB.pep:US-08-460-890A-62 + 63.00 131.30 7.33 100
/cgn2_6/prodata/2/iaa/5b_COMB.pep:US-08-167-641C-62 + 63.00 131.30 7.33 100

seq_name: /cgn2_6/prodata/2/iaa/5b_COMB.pep:US-09-245-041-11

seq_documentation block:

Sequence 11, Application US/09245041
Patent No. 6274339
GENERAL INFORMATION:
APPLICANT: Moore, K.
APPLICANT: Naigle, D.
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE DIAGNOSIS AND TREATMENT
TITLE OF INVENTION: OF BODY WEIGHT DISORDERS INCLUDING OBESITY
FILE REFERENCE: 7853-136
CURRENT APPLICATION NUMBER: US/09/245,041
CURRENT FILING DATE: 1999-02-05
EARLIER APPLICATION NUMBER: 60/093,630
EARLIER FILING DATE: 1998-07-21
EARLIER APPLICATION NUMBER: 60/104,978
EARLIER FILING DATE: 1998-10-20
NUMBER OF SEQ ID NOS: 131
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 11
LENGTH: 351
TYPE: PRT
ORGANISM: Mus musculus
FEATURE:
NAME/KEY: SITE
LOCATION: all xaa positions
OTHER INFORMATION: Xaa=unknown amino acid
US-09-245-041-11

alignment_scores:
Quality: 75.00 Length: 29
Ratio: 3.261 Gaps: 0
Percent Similarity: 79.310 Percent Identity: 51.724

alignment_block:
US-09-823-101-1 x US-09-245-041-11 ..
Align seg 1/1 to: US-09-245-041-11 from: 1 to: 351

433 AAGATACTAAAGAGAACTACCTCAGCTTAAACCAACCAAAAA 502
|||||
24 LysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 40

503 AAAAAAAAAAAGGGGGGGAACCAAGGGCA 539
|||||
40 LysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 52

seq_name: /cgn2_6/prodata/2/iaa/5b_COMB.pep:US-09-245-041-19

seq_documentation block:

Sequence 19, Application US/09245041
Patent No. 6274339
GENERAL INFORMATION:
APPLICANT: Moore, K.
APPLICANT: Naigle, D.
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE DIAGNOSIS AND TREATMENT
TITLE OF INVENTION: OF BODY WEIGHT DISORDERS INCLUDING OBESITY
FILE REFERENCE: 7853-136
CURRENT APPLICATION NUMBER: US/09/245,041
CURRENT FILING DATE: 1999-02-05
EARLIER APPLICATION NUMBER: 60/093,630
EARLIER FILING DATE: 1998-07-21
EARLIER APPLICATION NUMBER: 60/104,978
EARLIER FILING DATE: 1998-10-20
NUMBER OF SEQ ID NOS: 131
SOFTWARE: FastSeq for Windows Version 3.0

```

; SEQ ID NO 19
; LENGTH: 871
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-245-041-19

```

```

alignment_scores:
  Quality: 73.50      Length: 114
  Ratio: 1.246       Gaps: 9
  Percent Similarity: 51.754   Percent Identity: 29.825

```

```

alignment_block:
US-09-823-101-1 x US-09-245-041-19

```

```

Align seg 1/1 to: US-09-245-041-19 from: 1 to: 871

```

```

194 TGTGTAGTCACATCATGTCAGTCCACGAAATCCAGCTCCAGT 243
      |||||||
792 CysValGlyProLeuLeuGlnProAlaSerProAsnThrValGlnProLy 808
244 TTTCTATATGACGAGAAAGACAGAGAGAAATATGCCAAATAGTC. 292
      :: :: ::
808 sleuAsnLeuAlaGlu.....GlyLysSerPheCysProPheIleP 822
293 .....ACAGTGAATTTGTCTGAGTGTGAATTATGCGTGAATCTTGTGTC 337
      ||| |||
822 roHISThrSerIleMetGly.....Phe 829
338 TTTTATATTTCCAAATTTTATATAGACACAGAAATACTTTTATAGAAAT 387
      ||| :: ::
830 PheValAlaPheAsnAsnThrValLeu.....LysTyrLeu.....Ph 841
388 TCTGACTACA.....AAAATGTGTTTTTAATTCCTGTCGACAG 428
      ||| :: ::
841 eleuLeuSerPheGluIleLysAsnIleLeu.....CysCysSerV 855
429 TTCAGAAATATGCCATTAATGGAACAGATACTAAAGAGAAAGAACTACC 478
      ||| :: ::
855 alLysLys.....LysLysLys..... 860
479 TCAAGGTTAAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC 520
      ||| |||
861 .....LysLysLysLysLysLysLysLysLysLysLysLysLysLys 871

```

```

seq_name: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:US-08-477-451-15

```

```

seq_documentation_block:
; Sequence 15, Application US/08477451
; Patent No. 592865

```

```

; GENERAL INFORMATION:
; APPLICANT: Covacci, Antonello
; TITLE OF INVENTION: Helicobacter Pylori CagI Region
; NUMBER OF SEQUENCES: 46
; CORRESPONDENCE ADDRESS:
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608-2916
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/477,451
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: A35
; ATTORNEY/AGENT INFORMATION:
; NAME: McClung, Barbara G.
; REGISTRATION NUMBER: 33,113

```

```

; REFERENCE/DOCKET NUMBER: 0335.002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 510-601-2708
; TELEFAX: 510-655-3542
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1724 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-477-451-15

```

```

alignment_scores:
  Quality: 72.00      Length: 187
  Ratio: 0.986       Gaps: 9
  Percent Similarity: 39.037   Percent Identity: 22.460

```

```

alignment_block:
US-09-823-101-1/rev x US-08-477-451-15

```

```

Align seg 1/1 to: US-08-477-451-15 from: 1 to: 1724

```

```

418 GGAATTAACCAACCAATTTT..... 398
      |||
145 GlyMetGlnLysProLeuPheAlaGlnIleSerCysSerValLeuPheAl 1471
397 .....GTAGTCACA 388
1471 aleuPheCysArgSerSerAspArgTyrLeuArgLysTyrCysIleHis 1488
387 AATTCCTTAAGATATTCGTGCTGATTAATAATTTGGAATA..... 344
      |||
1488 erPheLeuIleIlePheTrpIleGlyTyrGlnAsnSerCysLeuLysPro 1504
343 .....TA 342
1505 ThrAsnThrIlePheValPheValArgAsnSerIlePheAspTrpSerPh 1521
341 AAAAGAACCAACATCAGCAT..... 320
      |||
1521 eSerGluTyrSerIleProHisHisLeuHisPheSerThrSerLysLysP 1538
319 ..AATTCACACACAGACAAATACCTGATATTTTGGCATATTTTCC 272
      |||
1538 roHISerProLeuSerIleSerHisSerTyrGlnSerProCysThrPro 1554
271 TCC.....TGTCCCTTTTCTGTCATATAGAAACCTGG.. 239
      |||
1555 SerValProSerLysAlaCysProPheGluGlnTyrLeuLeuLeuTrpse 1571
238 .AAGCTGGATTTTTCAGTGGACCTGACATGAGTGAGCTACACAGAA 190
      |||
1571 rGlnThrCysPhePheGlnAlaValSerAsnAsnLeuTyrTyrProIle 1588
189 TGTTCGTGTTTCATATTAAAGTCATGTGTCTGCTGCTGC..... 149
      |||
1588 euPheCys.....SerThrSplyCysTyrTrpAsn 1598
148 .....TGTCTGGCTACCTTTCCSCCATAGTACCTCTGTTTTC 108
      |||
1599 IleLeuGlnSerLeuAlaCysLysIleLysLeuIleGlySerLeuPheC 1615
107 TCTTTTTCSTA.....TTTGTCTAGCCCTTCC 79
      |||
1615 yspheValLeuHisLeuLysLysThrHisProPheGlnAlaLysProGlu 1631
78 CCTCACTTA 70
      |||
1632 ProTyrLeu 1634

```

```

seq_name: /cgn2_6/ptodata/2/1aa/5A_COMB.pep:US-08-221-817-13

```

```

seq_documentation_block:
; Sequence 13, Application US/08221817
; Patent No. 553151
; GENERAL INFORMATION:
; APPLICANT: Chantry, David
; APPLICANT: Gray, Patrick W.
; APPLICANT: Hoekstra, Merle F.
; TITLE OF INVENTION: A No. 5532151el G Protein-Coupled Receptor
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
; ADDRESSEE: Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/221,817
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/123,932
; FILING DATE: 17 SEP 1993
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5532151and, Greta E.
; REGISTRATION NUMBER: 35,302
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 474-6300
; TELEFAX: (312) 474-0448
; TELEX: 25-3856
; INFORMATION FOR SEO ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 576 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-221-817-13

alignment_scores:
Quality: 70.50 Length: 246
Ratio: 0.698 Gaps: 13
Percent Similarity: 41.057 Percent Identity: 21.545

alignment_block:
US-09-823-101-1 x US-08-221-817-13 ..
Align seg 1/1 to: US-08-221-817-13 from: 1 to: 576

42 ATGTGAACAAACAGAACCTCTAAGTTAAGTAGAGGGAGGCTAGAA 91
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
6 TlevalAlaasnThrValleuLeuLysAlaargLuglYglYglYs 22
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
92 CAAATAGAGAAAAGAGAAAAGAGAGCTATTAGTGGGGAAGGTAGG 141
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
22 nArgLysGly LysSerLysLysTrpArgGlnMetLeu..... 34
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
142 CCAGAGACAGACAGACAGACATGACCTTAATATGAAAC...AGAAC 188
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
35 .....GlnPheProHisIleSerGlnCysGlnGluLeuArgLe 47
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
189 ATTCTTGATGATGCTCACTCACTGCTCACTGCACAGAAAATCCAGCTCT 238
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
47 uSerLeuGlnArgAspTyrHisSerLeuCysGln...ArgHisArgIleG 63

```

```

239 CCAGTTTCTATATGACAGAG..... 258
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
63 LysArgLeuLeuPheArgGlnubhecysAlaThrArgProGluLeuSerArg 79
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
259 .....AAAAAGA 266
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
80 CysValAlaPheLeuAspGlyValAlaGluTyrGluValThrProAspAs 96
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
267 CAGAGAGGAAGATATGC.....CAAAAT.....AGTCACA 295
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
96 PLYsArgLysAlaCysGlyArgHisValThrGlnAsnPheLeuSerHisT 113
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
296 GT..... 297
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
113 hArgLysProAspLeuIleProGluValProArgGlnLeuValThrAsnGys 129
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
297 ..... 297
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
130 ThrGlnArgLeuGlnGlnGlyProCysLysAspLeuPheGlnGluLeuTh 146
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
298 .....GATTTGTCTGAGTGTGATATGCTGATCTTGTG 334
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
146 rArgLeuThrHisGluTyrLeuSerValAlaProPheAlaAspTyrLeu 162
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
335 TTCCTTTTATATTCACAAATTTTATAAGACACAGAAATACTTTTAAGA 384
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
163 AspSerIleTyrPheAsnArgPheLeuGln...TrrpLysTrpLeuGluAr 178
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
385 ATTTCTGACTACAAAATAATGTTT..... 409
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
178 gLInProValThrLysAsnThrPheArgGlnTyrArgValLeuGlyLysG 195
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
410 .....TTTAATTCCTGTGCCAGTTCAGAAATATGCCATATGSAACAA 454
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
195 LysGlyPheGlyGluValCysAlaCysGln..... 204
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
455 GATACTAAAGAGAAAGAACTACCTCAAGTTA.....AAAAACC 495
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
205 .....ValArgAlaThrGlyLysMetTyrAlaCysLysLysLe 217
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
496 AAAAAAAAAAAAAAAAAAAAAAGGGGGGGA 529
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
217 uGluLysLysArgIleLysLysArgLysGln 228

seq_name: /cgn2_6/prodata/2/iaa/5A_COMB.pep:US-08-454-439-13

seq_documentation_block:
; Sequence 13, Application US/08454439
; Patent No. 5591618
; GENERAL INFORMATION:
; APPLICANT: Chantry, David
; APPLICANT: Gray, Patrick W.
; APPLICANT: Hoekstra, Merle F.
; TITLE OF INVENTION: A No. 5591618el G Protein-Coupled Receptor
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
; ADDRESSEE: Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/454,439
; FILING DATE: 30-MAY-1995

```

```

CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/221,817
FILING DATE: 31-MAR-1994
APPLICATION NUMBER: 08/123,932
FILING DATE: 17 SEP 1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: No. 559161band, Greta E.
REGISTRATION NUMBER: 35,302
REFERENCE/DOCKET NUMBER: 31981
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312) 474-6300
TELEFAX: (312) 474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 576 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-454-439-13

```

```

alignment_scores:
Quality: 70.50 Length: 246
Ratio: 0.698 Gaps: 13
Percent Similarity: 41.057 Percent Identity: 21.545

```

alignment_block:

US-09-823-101-1 x US-08-454-439-13 ..

Align seg 1/1 to: US-08-454-439-13 from: 1 to: 576

```

42 ATGCTAACAGACAGCAAGCTCTAAGCTTAAGTGAGGGAAGCGCTGAA 91
   ::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|
6  ILeValAlaasnThrValLeuLeuLysAlaArgGluGlyGlyGlyLys 22
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
92 CAAATAGAGAAAGAGAAAAAGAGAGCTCATTAAGTGGGAAAGTAGG 141
   ||::|::|::|::|::|::|::|::|::|::|::|::|::|::|
22 nArgLysGly.LysSerLysLysThrArgLysMetLeu..... 34
142 CCAGAGAGCAGGCGAGAGCCACATGACCTTAATATGGAAAC..AGAC 188
   ||::|::|::|::|::|::|::|::|::|::|::|::|::|::|
35 .....GlnPheProHisLysSerGlnCysGluGluLeuArgLe 47
189 ATTCTTGTAAGTCACATGCTGCTGCGCACTGAAATCCAGCTCT 238
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|
47 uSerLeuGluArgAspTyrHisSerLeuCysGlu..ArgHisArgLleG 63
239 CCAAGTTTCTATATGCAGAG..... 258
63 LysArgLeuLeuPheArgGluPheCysAlaThrArgProGluLeuSerArg 79
259 .....AAAAAGA 266
80 CysValAlaPheLeuAspGlyValAlaGluTyrGluValThrProAspAs 96
267 CAGGAAGGAATATGC.....CAAAAT.....AGTCACA 295
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|
96 pLysArgLysAlaCysGlyArgHisValThrGlnAsnPheLeuSerHisT 113
296 GT..... 297
113 hArgLysProAspLeuIleProGluValProArgGlnLeuValThrAsnCys 129
297 ..... 297
130 ThrGlnArgLeuGluGlnGlyProCysLysAspLeuPheGlnGluLeuTh 146
298 .....GATTGTGTCGAGGCGCAATTAAGTGGTGTCTTGG 334
146 rArgLeuThrHisGluTyrLeuSerValAlaProPhe..AlaAspTyrLeu 162

```

```

335 TTCTTTTAAATTTCCAAATTTTATAGAGCACGAAATACCTTTTAGA 384
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
163 AspSerIleTyrPheAsnArgPheLeuGln...TriPylsThrPheuGluAr 178
385 ATTTCGACTACCAAAAAATTGGTTT..... 409
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
178 gGlnProValThrLysAsnThrPheArgGlnTyrArgValLeuGlyLysG 195
410 .....TTTAATTCCTGTCGCAAGTTCAGAAATATGCAATATGCAACAA 454
195 LysGlyPheGlyGluValCysAlaCysGln..... 204
455 GATTACTAAAGAAAGAAACACTACCTCAAGGTTA.....AAAAACC 495
205 .....ValArgAlaThrGlnGlyLysMetCysLysLysLysLysLys 217
496 AAAAAAAAAAAAAAAAAAAGGGGGGGGAA 529
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
217 uGluLysLysArgLleLysLysArgLysGlyGlu 228

```

seq_name: /cgn2_6/plodata/2/1aa/PCTUS_COMB.pep:PCT-US94-10487-13

```

seq_documentation_block:
; Sequence 13, Application PC/TUS9410487
; GENERAL INFORMATION:
; APPLICANT: ICOS Corporation
; TITLE OF INVENTION: A Novel G Protein-Coupled Receptor
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/10487
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/221,817
; FILING DATE: 31 MAR 1994
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/123,932
; FILING DATE: 17 SEP 1993
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Noland, Greta E.
; REGISTRATION NUMBER: 35,302
; REFERENCE/DOCKET NUMBER: 27866/31981
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 474-6300
; TELEFAX: (312) 474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 576 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US94-10487-13

```

alignment_scores:

Quality: 70.50 Length: 246
Ratio: 0.698 Gaps: 13
Percent Similarity: 41.057 Percent Identity: 21.545

Alignment_block:
US-09-823-101-1 x PCT-US94-10487-13 ..

Align seg 1/1 to: PCT-US94-10487-13 from: 1 to: 576

```

42 ATGCAAGACGACGACCTCTAGCTTAAGTGAAGGAGCGGTAGAA 91
   ::::::::::::::: |||||::: ||||| ||| 11
6 11eValAlaAsnThrValLeuLeuysAlaargGluGlyGlyGlyAs 22
92 CAAATAGAGAAAAAGAGAAAAAGAGTCATTAAGTGGGAGAGT 141
   ||: ||| |||::: ||||| |||::: ||:::
22 nArglySgLy.LySerlySlyStrArgGlnMetLeu..... 34
142 CCAGAGACGACGACAGACACATGACCTTAATATGAAAC..AGAAC 188
   ||| ||||| ::| ||::: |||
35 .....GlnPheProHisLeSerGlnCysGlnGlnLeuArgLe 47
189 ATTCTGTGTAGTCCACTGACGTCGACGTCGACGACGACGACGCT 238
   ||| |||::: |||::: |||::: |||:::
47 uSerLeuGluArgAspTyrHisSerLeuGln.....ArgHisArgIleG 63
239 CCAGTTTCTATATGACAGAG..... 258
   ::|::|::|::|::|
63 lYArgLeuLeuPheArgGlnPheCysAlaThrArgProGlnLeuSerArg 79
259 .....AAAAAGA 266
   ||
80 CysValAlaPheLeuAspGlyValAlaGluTyrGluValThrProAspAs 96
267 CAGSAGAGAAATATGC.....CAAAAT.....ACTGACA 295
   ||::: ||::: ||| |||||
96 pLysArgLysAlaCysGlyArgHisValThrGlnAsnPheLeuSerHisT 113
296 GT..... 297
   ::
113 hngLyProAspLeuLeuProGluValProArgGlnLeuValThrAsnGly 129
297 ..... 297
130 ThrGlnArgLeuGlnGlnGlyProCysLysAspLeuPheGlnLeuTh 146
298 .....GATTTGTCTGAGTGGTGAATATGCTGATCTTGG 334
   ::::: ||::: ||::: ||| ||| |||
146 rArgLeuThrHisGluTyrLeuSerValAlaProPhe..AlaAspTyrLeu 162
335 TTCTTTTATATTTCCAAATTTTATAAGACACAGAAATACTTTAAGA 384
   ::|||::: |||::: |||:::
163 AspSerIleTyrPheAsnArgPheLeuGln..TrpLysTrpLeuGlnArg 178
365 ATTCTGACTCAAAATAATGCTTT..... 409
   ::||| |||
178 gGlnProValThrLysAsnThrPheArgGlnTyrArgValLeuGlyLysG 195
410 ....TTTAATTCCTGTGCCAAGTTCAGAAATATGCCATATGGAACAA 454
   |||::: |||
195 lYgLyPheGlyGlyValAlaCysAlaCysGln..... 204
455 GATACTAAAGAGAAAGAACTACCTCAAGGTAA.....AAAAACC 495
   ::::: |||::: |||
205 .....ValArgAlaThrGlnLysMetTyrAlaCysLysLysLe 217
496 AAAAAAAAAAAAAAAAAAGGGGGGGA 529
   ::|||::: ||| |||
217 uGluLysLysArgIleLysLysArgLysGlyGlu 228

```

seq_name: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-08-942-686-2

seq_documentation_block:
; Sequence 2, Application US/08942686

```

; Patent No. 6183988
; GENERAL INFORMATION:
; APPLICANT: Bloch, Donald B.
; APPLICANT: Bloch, Kenneth D.
; TITLE OF INVENTION: LEUCOCYTE-SPECIFIC PROTEIN AND GENE, AND
; TITLE OF INVENTION: METHODS OF USE THEREOF
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN AND FOX, P.L.L.C.
; STREET: 1100 NEW YORK AVENUE, N.W. SUITE 600
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/942,686
; FILING DATE: 02-OCT-1997
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/027,347
; FILING DATE: 02-OCT-1996
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: GOLDSTEIN, JORGE A.
; REGISTRATION NUMBER: 29,021
; REFERENCE//DOCKET NUMBER: 0609.4330001/JAG/BJD
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-2600
; TELEFAX: 202-371-2540
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 753 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-942-686-2

```

alignment_scores:
Quality: 69.50 Length: 91
Ratio: 1.311 Gaps: 4
Percent Similarity: 58.242 Percent Identity: 26.374

alignment_block:
US-09-823-101-1 x US-08-942-686-2 ..

Align seg 1/1 to: US-08-942-686-2 from: 1 to: 753

```

14 ACCAGAGATCTGAGTGAAGACTCACATTATGGTAAACAGACGACCTCC 63
   ||| |||::: |||::: |||
369 ThrValAlaPheAlaAsnAsnSerThrLeuGlyLys.....Pr 381
64 TAGAGTTAAGTCAGCGGAAGGCTAGAACAAATATGAAAAAGAAAAA 113
   ||| ||| ||| ||| ||| |||
381 Olys.....ArgLysArgArgLysL 388
114 AGAGAGTCATTAGTGGGAAAGTAGCCAGAGAGACGACGACACAC 163
   ||||| |||::: |||::: |||
388 ysArgGlyHisGlyTyrPserArgMetArgMetArgGlnLysAsnSer 404
164 ATGACCTTAATATGGAAGAACAGACATCTTGTGTAGCTCCATCATGCT 213
   ::|::: |||::: |||
405 Gln.....GlnAsnAspAsnSerLysAlaAspGlyGlnValVa 417
214 CAGTGCACCTGAAGAAATCCAGCTCCAGTTTCTATATGACGAGAAAAA 263
   |||::: |||::: |||
417 lSerSerGluLysLysAlaAsnValAlaAsnLeuLysAspLeuSerLysIleA 434

```



```

63  LyaArgLeuPheArgGluPheCysAlaThrArgProGluLeuSerArg 79
259 .....AAAAAGCA 266
80  CysValAlaPheLeuAspGlyValAlaGluTyrGluValThrProAspAs 96
267  CAGGAGAGAAATATGC.....CAAAAT.....AGTCACA 295
96  pLysArgLysAlaCysGlyArgHisValThrGlnAsnPheLeuSerHisT 113
296  GT..... 297
113  hGlyProAspLeuLeuProGluValProArgGlnLeuValThrAsnCys 129
297 ..... 297
130  ThrGlnArgLeuGlnGlnGlyProCysLysAspLeuPheGlnGluLeuTh 146
298 .....GATTTGTCTGAGTGTGAATATGCGTATCTTTG 334
146  rArgLeuThrHisGluTyrLeuSerValAlaProPheAlaAspTyrLeu 162
335  TTCTTTTATTTCCAAATTTTATAGACACAGAAATCTTTTACGA 384
163  AspSerIleTyrPheAsnArgPheLeuGln...TriLysTrpLeuGlnAr 178
385  ATTTCGACTACAAAAATTTGTTT..... 409
178  gGlnProValThrLysAsnThrPheArgGlnTyrArgValLeuGlyLysG 195
410  ....TTTAATTCCTGTGCCAAGTTCAGAAATATGCCATTAAGCAAA 454
195  LylGlyPheGlyGluValCysAlaCysGln..... 204
455  GATTACTAAAGAGAAACTACCTCAAGCTTA.....AAAAACC 495
205  .....ValArgAlaThrGlyLysMetTyrAlaCysLysLysLe 217
496  AAAAAAAAAAAAAAAAAAAAAAGGGGGGGA 529
217  uGluLysLysArgLleLysLysArgLysGlyLys 228

seq_name: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:US-08-477-451-3

seq documentation block:
: Sequence 3, Application US/08477451
: Patent No. 592865
: GENERAL INFORMATION:
: APPLICANT: Covaccl, Antonello
: TITLE OF INVENTION: Helicobacter Pylori CagI Region
: NUMBER OF SEQUENCES: 46
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Chiron Corporation
: STREET: 4560 Horton Street
: CITY: Emeryville
: STATE: CA
: COUNTRY: USA
: ZIP: 94608-2916
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: IBM PC compatible
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/477,451
: FILING DATE: 07-JUN-1995
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: McClung, Barbara G.
: REGISTRATION NUMBER: 33,113
: REFERENCE/DOCKET NUMBER: 0335, 002
: TELECOMMUNICATION INFORMATION:

```

```

: TELEPHONE: 510-601-2708
: TELEFAX: 510-655-3542
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 3174 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-477-451-3

alignment_scores:
  Quality: 68.50      Length: 120
  Ratio: 1.557      Gaps: 4
  Percent Similarity: 36.667      Percent Identity: 24.167

alignment_block:
  US-09-823-101-1/rev x US-08-477-451-3 ..

Align seg 1/1 to: US-08-477-451-3 from: 1 to: 3174

537  CCCCTTTGTTCCCCCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 488
1100  ProIleGlySerValPhePhePheLeuCysPhePheGlyPhePh 1116
487  TAACCTTGA..... 479
1116  eValIleIleCysProIleArgThrGlnAsnAsnGlnThrGlnAsnPhG 1133
478  ....GGTAGTTTCTTCTTTTAACTTATC..... 455
1133  InLysLysAlaSerPheLeuIleAlaIleIleIleValAspGlnLeu 1149
454  ....TTGTTTCAT..... 446
1150  SerLeuLeuHisProArgGlnTyrSerGlnThrCysSerAsnArgTyrAr 1166
445  ....TATGGCATATTTCTGAACTTGGCACAGGAAATTAATAAAC 406
1166  gLeuProProLeuThrLeuThrPValAlaArgLysGlnLysGlnSerPro 1183
405  CAATTTTGTAGTCAGAAATTTCTTAAAGTATTTCTGTGCTCTATATA 356
1183  sPhePheCysHisGlnArgPheSerLysTyrProLeuCys..... 1196
355  AATTGGAAATATTAAGAACAAAGAAATCAGCCATTAATTCACACTCAG 306
1197  .....LeuProLeuLys 1200
305  ACAAAATCAC 296
1200  sArgAsnHis 1203

seq_name: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:US-08-748-947A-2

seq documentation block:
: Sequence 2, Application US/08748947A
: Patent No. 5854031
: GENERAL INFORMATION:
: APPLICANT: Jigami et al., Yoshifumi
: TITLE OF INVENTION: MANNOSE-1-PHOSPHATE TRANSFERASE GENE
: TITLE OF INVENTION: FROM YEAST, AND A PROCESS FOR PRODUCING
: TITLE OF INVENTION: MANNOSE-1-PHOSPHATE-CONTAINING ACIDIC SUGAR CHAINS AND
: TITLE OF INVENTION: PHOSPHATE-CONTAINING ACIDIC SUGAR CHAINS BY USE OF THE
: NUMBER OF SEQUENCES: 2
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Fish & Richardson P.C.
: STREET: 601 Thirteenth Street, NW
: CITY: Washington
: STATE: DC
: COUNTRY: USA

```

```

; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/748,947A
; FILING DATE: 14-NOV-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 299509/1995
; FILING DATE: 17-NOV-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Ellison, Eldora L.
; REGISTRATION NUMBER: 39,967
; REFERENCE/DOCKET NUMBER: 08206/006001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202/783-5070
; TELEFAX: 202/783-2331
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 446 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-748-947A-2

```

```

alignment_scores:
  Quality: 67.00      Length: 123
  Ratio: 1.218      Gaps: 6
Percent Similarity: 44.715      Percent Identity: 24.390

```

```

alignment_block:
US-09-823-101-1/rev x US-08-748-947A-2

```

```

Align seq 1/1 to: US-08-748-947A-2 from: 1 to: 446

```

```

353 TTTCGAATATTAAGAACAAAGACAGCATTAATTCACCACTCAGAC 304
||||:|||||:||||:|||||
324 PheSnaSncysGluPheThSerAsn..... 332
303 AAATCAGCTGAGCTATTGTCATATTTCCTGCTCTTTTC.... 258
||||:|||||:||||:|||||
333 .....PheGluIleGlyAsnLeuAsnPhenylrgrs 343
257 ..TCTGATATAGAAACTG.....GAGAGCTGGAATT 228
||||:|||||:||||:|||||
343 ePrOlaIaTyArGlySpherheAsnTyrlleAspGluGlyGlyIle 359
227 TTTCAGTGGCAGCTGAGCATGATGAGCTACACAGAATGTCGTTC 178
||||:|||||:||||:|||||
360 PheTyItrPtyStrSer.....AspSe 367
177 CATATTAAGTCATGTCGTCCTGCTCTGCTGCTGCTCTTCCA 128
||||:|||||:||||:|||||
367 rIleIleIleThrlleGlyLeuSerMetIle.....LeuPrO 380
127 CTAATGACCTCTCTTTTCTCTTTTCTATTTGTTCTAGCCCTTCC 78
||||:|||||:||||:|||||
380 yAsrPlsIleIleSpherheGluSnaIleGlyPhe..... 391
77 CTCACCTTAACCTTAGAGTCTGCTGTTGTTACATATGAGCTTCA 28
||||:|||||:||||:|||||
392 .....HisTyAspLysTyAsnAsnSncysProLeuAs 402
27 CTCAGATCTCTGCTTCCA 9
||||:|||||:||||:|||||
402 nAspAspIleIleTyAsnGln 408

```

```

seq_name: /cgn2_6/ptodata/2/1aa/6A.COMB.pep:US-09-068-650-5

```

```

seq_documentation_block:
; Sequence 5, Application US/09068650
; Patent No. 6083720
; GENERAL INFORMATION:
; APPLICANT: Chirodoczek Jadwiga
; APPLICANT: Fender Pascal
; TITLE OF INVENTION: Dodecahedral Adenoviral Protein Complex,
; FILE REFERENCE: 3339-368
; CURRENT APPLICATION NUMBER: US/09/068,650
; CURRENT FILING DATE: 1998-07-31
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 40
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: hybrid peptide
US-09-068-650-5

```

```

alignment_scores:
  Quality: 66.00      Length: 23
  Ratio: 3.667      Gaps: 0
Percent Similarity: 78.261      Percent Identity: 60.870

```

```

alignment_block:
US-09-823-101-1 x US-09-068-650-5

```

```

Align seq 1/1 to: US-09-068-650-5 from: 1 to: 40

```

```

452 CAAGATTAAGTAAAGAAAGAACTCAAGCTTAATAAAACCAAAAA 501
:::|||||:|||||:|||||:|||||
17 GluSpGluSerLysLysLysLysLysLysLysLysLysLysLys 33
502 AAAAAAAAAAAAAAAAAAG 520
|||||:|||||:|||||:|||||
33 slyLysLysLysLysLysLys 39

```

```

seq_name: /cgn2_6/ptodata/2/1aa/6B.COMB.pep:US-09-612-126-11

```

```

seq_documentation_block:
; Sequence 11, Application US/09612126
; Patent No. 6284726
; GENERAL INFORMATION:
; APPLICANT: TEMPLE UNIVERSITY - OF THE COMMONWEALTH SYSTEM OF
; TITLE OF INVENTION: INHIBITION OF ANGIOGENESIS BY PEPTIDE ANALOGS OF HIGH
; FILE REFERENCE: 6056-258 CT1
; CURRENT APPLICATION NUMBER: US/09/612,126
; CURRENT FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: 60/107,844
; PRIOR FILING DATE: 1998-11-10
; PRIOR APPLICATION NUMBER: PCT/US99/26377
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 11
; LENGTH: 179
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Human high
; OTHER INFORMATION: molecular weight kininogen light chain amino acids
; OTHER INFORMATION: Glu(448) through Ser(626)
US-09-612-126-11

```

```

alignment_scores:
  Quality: 66.00      Length: 68
  Ratio: 1.737      Gaps: 4
Percent Similarity: 55.882      Percent Identity: 32.353

```

alignment_block:

US-09-823-101-1 x US-09-612-126-11 ..

Align seg 1/1 to: US-09-612-126-11 from: 1 to: 179

```
77 GGGGAAGGCTAGACAAATAGGAAAAAGAGAGAGCTCATTA 126
   |||:| |||:| |||:| |||:| |||:| |||:| |||:|
23 GLYHISValleuAspHisGlyHisLys.....HisLysHisGlyHisGcl 37
127 GTGGGGAAG.....GTAGCCAGAGAGCAGCAGACACCATGA 167
   ||| ||| |||:| |||:| |||:| |||:| |||:| |||:|
37 yHisGlyLysHisLysAsnLysGlyLysAsnGlyLysHisAsnGly. 53
168 CCTTAATATGAAACAGAACATTTGTGTAGGTCACATCATCTCACT 217
   ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
54 .....TrpLysThrGluHis.....LeuAlaSer 61
218 GCCACTGAAATAATCCAGCTCTCCAGTTTCTATATGACAGAGAAAAAGAC 267
   |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:|
62 SerSerGluAspSerThrThrProSerAlaGlnThrGlnGlyLysThrGcl 78
268 AGGA 271
   |||
78 uGly 79
```

seq_name: /cgn2_6/ptodata/2/1aa/6B_COMB.pep:US-09-612-126-8

seq_documentation_block:

```
; Sequence 8, Application US/09612126
; Patent No. 6284726
; GENERAL INFORMATION:
; APPLICANT: TEMPLE UNIVERSITY - OF THE COMMONWEALTH SYSTEM OF
; TITLE OF INVENTION: INHIBITION OF ANGIOGENESIS BY PEPTIDE ANALOGS OF HIGH
; TITLE OF INVENTION: MOLECULAR WEIGHT KININOGEN DOMAIN 5
; FILE REFERENCE: 6056-258 CT1
; CURRENT APPLICATION NUMBER: US/09/612,126
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: 60/107,844
; PRIOR FILING DATE: 1998-11-10
; PRIOR APPLICATION NUMBER: PCT/US99/26377
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Patentln Ver. 2.1
; SEQ ID NO 8
; LENGTH: 186
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Human high
; OTHER INFORMATION: molecular weight kininogen light chain amino acids
; OTHER INFORMATION: His(441) through Ser(626)
US-09-612-126-8
```

alignment_scores:

Quality:	66.00	Length:	68
Ratio:	1.737	Gaps:	4
Percent Similarity:	55.882	Percent Identity:	32.353

alignment_block:

US-09-823-101-1 x US-09-612-126-8 ..

Align seg 1/1 to: US-09-612-126-8 from: 1 to: 186

```
77 GGGGAAGGCTAGACAAATAGGAAAAAGAGAGAGCTCATTA 126
   |||:| |||:| |||:| |||:| |||:| |||:| |||:|
30 GLYHISValleuAspHisGlyHisLys.....HisLysHisGlyHisGcl 44
127 GTGGGGAAG.....GTAGCCAGAGAGCAGCAGACACCATGA 167
   ||| ||| |||:| |||:| |||:| |||:| |||:| |||:|
44 yHisGlyLysHisLysAsnLysGlyLysAsnGlyLysHisAsnGly. 60
```

168 CCTTAATGAAACAGAACATTTGTGTAGGTCACATCATCTCACT 217

||| ||| ||| ||| ||| ||| ||| ||| ||| |||

61TrpLysThrGluHis.....LeuAlaSer 68

218 GCCACTGAAATAATCCAGCTCTCCAGTTTCTATATGACAGAGAAAAAGAC 267

|||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:|

69 SerSerGluAspSerThrThrProSerAlaGlnThrGlnGlyLysThrGcl 85

268 AGGA 271

|||

85 uGly 86

seq_name: /cgn2_6/ptodata/2/1aa/6B_COMB.pep:US-09-612-126-1

seq_documentation_block:

```
; Sequence 1, Application US/09612126
; Patent No. 6284726
; GENERAL INFORMATION:
; APPLICANT: TEMPLE UNIVERSITY - OF THE COMMONWEALTH SYSTEM OF
; TITLE OF INVENTION: INHIBITION OF ANGIOGENESIS BY PEPTIDE ANALOGS OF HIGH
; TITLE OF INVENTION: MOLECULAR WEIGHT KININOGEN DOMAIN 5
; FILE REFERENCE: 6056-258 CT1
; CURRENT APPLICATION NUMBER: US/09/612,126
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: 60/107,844
; PRIOR FILING DATE: 1998-11-10
; PRIOR APPLICATION NUMBER: PCT/US99/26377
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Patentln Ver. 2.1
; SEQ ID NO 1
; LENGTH: 255
; TYPE: PRT
; ORGANISM: Human
US-09-612-126-1
```

alignment_scores:

Quality:	66.00	Length:	68
Ratio:	1.737	Gaps:	4
Percent Similarity:	55.882	Percent Identity:	32.353

alignment_block:

US-09-823-101-1 x US-09-612-126-1 ..

Align seg 1/1 to: US-09-612-126-1 from: 1 to: 255

```
77 GGGGAAGGCTAGACAAATAGGAAAAAGAGAGAGCTCATTA 126
   |||:| |||:| |||:| |||:| |||:| |||:| |||:|
99 GLYHISValleuAspHisGlyHisLys.....HisLysHisGlyHisGcl 113
127 GTGGGGAAG.....GTAGCCAGAGAGCAGCAGACACCATGA 167
   ||| ||| |||:| |||:| |||:| |||:| |||:| |||:|
113 yHisGlyLysHisLysAsnLysGlyLysAsnGlyLysHisAsnGly. 129
168 CCTTAATGAAACAGAACATTTGTGTAGGTCACATCATCTCACT 217
   ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
130 .....TrpLysThrGluHis.....LeuAlaSer 137
218 GCCACTGAAATAATCCAGCTCTCCAGTTTCTATATGACAGAGAAAAAGAC 267
   |||:| |||:| |||:| |||:| |||:| |||:| |||:| |||:|
138 SerSerGluAspSerThrThrProSerAlaGlnThrGlnGlyLysThrGcl 154
268 AGGA 271
   |||
154 uGly 155
```

